(6)
13.8.70

Psychological Review

EDITED BY

THEODORE M. NEWCOMB, UNIVERSITY OF MICHIGAN

VOLUME 61, 1954

PUBLISHED BIMONTHLY

BY THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
PRINCE AND LEMON STS., LANCASTER, PA.
AND 1333 SIXTEENTH ST. N. W., WASHINGTON 6, D. C.

Entered as second-class matter July 13, 1897, at the post-office at Lancaster, Pa., under Act of Congress of March 3, 1879

Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized Jan. 8, 1948

Send all business communications, including address changes, to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must arrive by the 25th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

LANCASTER PRESS, INC., LANCASTER, PA.

CONTENTS OF VOLUME 61

David Katz

VOL. 61, NO. 1 JANUARY, 1954

THE PSYCHOLOGICAL REVIEW

DAVID KATZ

1884–1953

David Katz, Professor Emeritus at the University of Stockholm, died of a sudden heart attack on February 2, 1953. By those who attended the International Congress of Psychology in Stockholm in July, 1951, he will be remembered as the indefatigable organizer and genial host of the congress. In the history of psychology his name will be associated with significant contributions to almost every field of psychology, pure and applied; and he will be cited as one of this century's outstanding exponents of psychological phenomenology. In the memories of those who knew him and loved him he will live as a gentle, humble man, persistently curious about everything that had to do with human nature, brilliant in his intuitions, tireless in his research, unfailingly generous and courteous in controversy.

Katz was born in Kassel, Germany, on October 1, 1884. His early education was in Kassel, his university education in Berlin, Munich, and Göttingen, where he received his doctoral degree in 1906. In Göttingen he was one of G. E. Müller's most brilliant pupils. Later he became Müller's assistant, and, in 1911, Privat Dozent. During World War I he was called to army service for four years, returning afterwards to his post in Göttingen. It was during his Göttingen period that he completed his now classic researches on the experimental phenomenology of color, and began his less well-known but equally significant work on touch. In 1919 he accepted the chair of psychology and education at the University of Rostock, where he developed what eventually became one of the most productive psychological laboratories in Europe. In 1933 the National Socialist party came into power, and Katz, as a non-Aryan, was deprived of his position. Fortunately his British friends were willing to provide hospitality and, for the next four years, first in Manchester and later in London, he was able to pursue his scientific work. In 1937 he accepted the chair of education (including psychology) at the University of Stockholm, where he remained until his retirement in 1952.

Katz paid two visits to the United States, in 1929 as Visiting Professor at the University of Maine and in 1950 as Hitchcock Lecturer at the University of California.

During his period as G. E. Müller's assistant, Katz was fond of relating, an attractive young Russian girl was admitted as a student. In reply to Katz's query, Müller characterized her as "eine Madonna mit einer Bombe." Rosa Heine did not blow up the Institute, thereby failing to conform to Müller's stereotype of the Russian, but she speedily conquered Müller's assistant. Katz and Rosa Heine were married in 1919. Numerous joint publications attest to their productivity as a scientific team. Their two sons, now launched on their own professional careers, were made prematurely famous by their parents' book, *Gespräche mit Kindern* (1927).

To review Katz's contributions to psychology would be a major undertaking. As a scientist he had "green fingers." He had but to touch a problem, and it readily blossomed and bore fruit. His list of publications includes more than 100 titles, of which at least 20 are substantial books and monographs. Among these one finds contributions to animal, child, educational, abnormal, and social psychology, to the experimental psychology of perception, motivation, learning, and thinking, to systematic theory, and to laboratory instrumentation. It may be that he scattered his energies too widely; certainly, not all his researches are of equal merit. It was his genius, however, to find in the commonplace observations of daily life problems which, when viewed in a larger context, became significant, and to make psychological capital out of every new experience with which good or bad fortune provided him. Thus, his wartime assignment to a military hospital led to a pioneer study of the psychological problems of amputees, and later to the invention of a device for the training of students in the technique of percussion; the feeding problems of his children contributed to his interest in constitutional typology and in the theory of hunger and appetite; his own difficulty with the English and Swedish languages challenged him as a psychologist to do some experiments on problems of language and thinking.

It was also his genius to find simple and inexpensive ways of attacking major problems. Katz belonged perforce to the cardboard and thumbtack school; but he never allowed a meager budget to hamper his activity. In Rostock he was faced with the task of developing a research institute on an annual budget of approximately $125. Some of his problems required the use of animals. He could not afford a regular animal laboratory; so he bought some chicks. Out of his chicken yard came the well-known *Hackgesetz*, the studies of chickens reared in isolation, the studies of "counting" behavior in chickens, and the experiments that led to the "avidity" theory of appetite. While in England, lacking an adequate laboratory, he pursued his tactual researches by undertaking some assignments for the flour millers, who were concerned about the elasticity of their dough. When he arrived in Sweden, he was assigned a small apartment as a laboratory. The kitchen promptly became a workshop, the bathroom became a photographic darkroom, a fifteen-year-old boy served as technician, and with cardboard, thumbtacks, bathroom scales, and sticks of wood, the laboratory began to produce research. When one thinks of David Katz, one wonders sometimes whether handsome budgets are a hindrance or an aid to productivity.

The frustrated graduate student in search of a doctoral problem has but to thumb through a few of Katz's publications to find a wealth of inviting questions and challenging hypotheses that will draw him straight to the laboratory. The human hand as a unitary sense organ analogous to the eye, the composite photograph as a device for the study of group characteristics, the sensory basis of the phenomenon of elasticity, the phantom limb of the amputee, the ability of certain deaf people to appreciate music, and a host of other apparent byways of psychological investigation were opened up by Katz and redirected towards the central problem. It was characteristic of his restless curiosity, however, that he was frequently content to blaze the trail, bequeathing to another generation the task of exploiting it.

The unity within Katz's apparent diversity of interest is to be found in his consistent application of the phenomenological method. He was interested in the prediction and control of behavior, in the social and biological determinants of behavior, in the tricky problems of instrumentation, in the broader problems of psychological theory, but behind it all was a persistent, a passionate curiosity about the world of phenomena. For Katz the most fascinating thing to wonder about was a human experience. It might be a simple color or sound, or the strange beauty of an El Greco picture, or the peculiar sensations that accompany the crunching of a nut between the teeth, or the ineffable satisfyingness of a cool draught of beer on a warm day. All experience was something to appreciate and to wonder about. For him the first task of the psychologist—not really a task, but a pleasure—was to observe and describe without bias both the salient characteristics and the subtle nuances of ordinary human experience. Phenomenology for him was essentially an attitude of "disciplined naiveté." From descriptive analysis one proceeds to experiment and to theory, but no psychological theory, he argued, could be complete if it excluded any of the essential variables of human experience.

Katz's psychological phenomenology is best exemplified in his studies of color and touch, *Die Erscheinungsweisen der Farben* (1911 [1]) and *Der Aufbau der Tastwelt* (1925). Influenced by the physiologist Hering and the philosopher Husserl he insisted that the psychologist should begin by deliberately "bracketing" his physical, physiological, and philosophical biases and attempt to observe phenomena as they are actually presented. The phenomenal world thus viewed contains properties and relationships that escape the notice of the physically or physiologically oriented observer. The classical psychologist was content to order colors in terms of hue, brightness, and saturation; Katz saw them also varying in mode of appearance, pronouncedness, insistence, transparency, inherence, and stability. Classical psychology was busily mapping the patterns of pressure, pain, warm, and cold spots on the skin, and searching for receptors; Katz went further, and explored the active process of "touching" (*tasten*), discovering here, too, modes of appearance, properties of organization, and unsuspected kinds of sensitivity. It is unfortunate that, while his visual studies have been widely appreciated, his richly suggestive book on the world of touch has received relatively little notice.

During recent years the word *phenomenological* has tended to expand its meaning. It is coming to suggest an easy-going, intuitive, sympathetic "seeing the world as the other fellow sees it," an approach that permits one to take things at their face value and to avoid the rigors of experimentation and theory construction. This is definitely not the kind of psychological phenomenology that Katz advocated. True, he was interested in the "fuzzy" aspects of experience; but for him the "fuzziness" of a phenomenon was no excuse for careless observation or undisciplined thinking. Good phenomenology, he held, requires at least as much training and discipline as does good Titchenerian introspection. Nor does phenomenology lead away from experimentation and theory; it is an essential first step in the direction of more imaginative experimentation and sounder theory.

Katz adhered to no "school" of psychology, nor—which is strange in a German of his generation—did he ever attempt to found a school. In his sympathies he stood closest to the Gestalt theorists; indeed, his pioneer work on

[1] Later revised as *Der Aufbau der Farbwelt* (1930); abridged and translated into English as *The World of Colour* (1935).

phenomenal constancy must be regarded as basic to the Gestalt theory of perception, and his more recent experiments on thinking belong in the Gestalt tradition. His interests were too varied, however, to fit neatly within any formal system, and we find him in his *Gestaltpsychologie* (1944) expressing impatience with the narrowness of the Gestalt approach. Like Stern he believed that every part process must be understood ultimately in terms of the total person, but he lacked Stern's compulsion to turn his personalism into a philosophy. With Jaensch he shared an interest in the possibilities of typology, but for him typology was a problem for research rather than a revelation. He found merit in the developmental approach, both ontogenetic and phylogenetic, but he rejected the extremes of both nativism and empiricism. He was willing to accept physiological evidence and to do physiological experiments when he felt that such would help to clarify a psychological problem, but he refused to accord to physiological constructs any unique explanatory value.

It is perhaps best to think of Katz as essentially a pioneer, catholic rather than eclectic, ready to adapt to his purposes any tool, material or conceptual, that looks useful, but never forgetting the purpose for which he has selected it. For Katz there was a single purpose that persisted throughout his scientific life. It was, to put it in old-fashioned language, to understand the phenomena of the human mind. Those who see as a challenge to science all the phenomena of human mentality will find in Katz a kindred spirit.

ROBERT B. MACLEOD

Cornell University

Psychological Review
Vol. 61, No. 1, 1954

THE PHYSIOLOGY OF MOTIVATION

ELIOT STELLAR

The Johns Hopkins University

In the last twenty years motivation has become a central concept in psychology. Indeed, it is fair to say that today it is one of the basic ingredients of most modern theories of learning, personality, and social behavior. There is one stumbling-block in this noteworthy development, however, for the particular conception of motivation which most psychologists employ is based upon the outmoded model implied by Cannon in his classical statement of the local theories of hunger and thirst (23). Cannon's theories were good in their day, but the new facts available on the physiological basis of motivation demand that we abandon the older conceptualizations and follow new theories, not only in the study of motivation itself, but also in the application of motivational concepts to other areas of psychology.

This argument for a new theory of motivation has been made before by Lashley (42) and Morgan (47). But it is more impelling than ever today because so much of the recent evidence is beginning to fit into the general theoretical framework which these men suggested. Both Lashley and Morgan pointed out that the local factors proposed by Cannon (e.g., stomach contractions or dryness of the throat) are not necessary conditions for the arousal of motivated behavior. Instead, they offered the more inclusive view that a number of sensory, chemical, and neural factors cooperate in a complicated physiological mechanism that regulates motivation. The crux of their theory was described most recently by Morgan as a *central motive state* (*c.m.s.*) built up in the organism by the combined influences of the sensory, humoral, and neural factors. Presumably, the amount of motivated behavior is determined by the level of the *c.m.s.*

Beach (8, 11), in his extensive work on the specific case of sexual motivation, has amply supported the views of Lashley and Morgan. But the important question still remains: Do other kinds of motivated behavior fit the same general theory? As you will see shortly, a review of the literature makes it clear that they do. As a matter of fact, there is enough evidence today to confirm and extend the views of Lashley, Morgan, and Beach and to propose, in some detail, a more complete physiological theory of motivation.

There are a number of ways to present a theoretical physiological mechanism like the one offered here. Perhaps the best approach is to start with an overview and summarize, in a schematic way, the major factors at work in the mechanism. Then we can fill in the details by reviewing the literature relevant to the operation of each factor. Some advantage is lost by not taking up the literature according to behavioral topics, that is, different kinds of motivation. But the procedure adopted here lets us focus attention directly on the theory itself and permits us to make some very useful comparisons among the various kinds of motivation. Once the theoretical mechanism and the evidence bearing on it are presented, the final step will be to evaluate the theory and show what experiments must be done to check it and extend it.

THEORETICAL SCHEME

A schematic diagram of the physiological mechanism believed to be in control of motivated behavior is shown in Fig. 1. The basic assumption in this scheme is that *the amount of motivated behavior is a direct function of the amount of activity in certain excitatory centers of the hypothalamus.* The activity of these excitatory centers, in turn, is determined by a large number of factors which can be grouped in four general classes: (*a*) *inhibitory hypothalamic centers* which serve only to depress the activity of the excitatory centers, (*b*) *sensory stimuli* which control hypothalamic activity through the afferent impulses they can set up, (*c*) *the internal environment* which can influence the hypothalamus through its rich vascular supply and the cerebrospinal fluid, and (*d*) *cortical and thalamic centers* which can exert excitatory and inhibitory influences on the hypothalamus.

As can be seen, the present theory holds that the hypothalamus is the seat of Morgan's *c.m.s.* and is the "central nervous mechanism" Lashley claimed was responsible for "drive." Identifying the hypothalamus as the main integrating mechanism in motivation makes the experimental problem we face more specific and more concrete than ever before. But it also makes it more complicated, for the physiological control of the hypothalamus is exceedingly complex. The influence of the internal environment on the hypothalamus is changing continuously according to natural physiological cycles, and of course it may often be changed directly by the chemical and physical consequences of consummatory behavior (see Fig. 1). Sensory stimuli may also have varied effects on the hypothalamic mechanism, depending upon their particular pattern, previous stimulation, previous learning, sensory feedback from the consummatory behavior itself, and the influence the internal environment has already exerted on the hypothalamus. Similarly, the influence of the cortex and thalamus will add to the hypothalamic activity already produced by sensory stimuli and the internal environment. Presumably, these cortical and thalamic influences may result directly or indirectly from sensory stimulation, but they may also be controlled partly by the "upward drive" of the hypothalamus itself (43). Then, to complicate the picture even more, there are the inhibitory centers of the hypothalamus which are also controlled by the various internal changes, sensory stimuli, and cortical and thalamic influences. These centers, presumably, depress the activity of the excitatory centers and, therefore, attenuate their output.

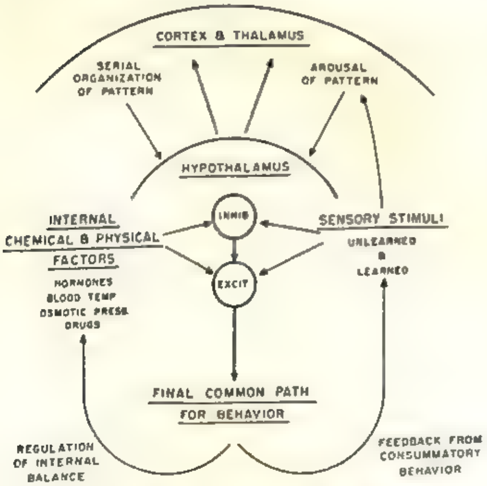
CORTEX & THALAMUS
SERIAL ORGANIZATION OF PATTERN
AROUSAL OF PATTERN
HYPOTHALAMUS
INHIB
EXCIT
INTERNAL CHEMICAL & PHYSICAL FACTORS
HORMONES
BLOOD TEMP
OSMOTIC PRESS
DRUGS
SENSORY STIMULI
UNLEARNED & LEARNED
FINAL COMMON PATH FOR BEHAVIOR
REGULATION OF INTERNAL BALANCE
FEEDBACK FROM CONSUMMATORY BEHAVIOR

FIG. 1. Scheme of the physiological factors contributing to the control of motivated behavior. (See text.)

Fortunately, this mechanism is not as formidable against experimental attack as it might appear. The basic experimental approach is to isolate the controlling factors in any type of motivation and determine their relative contributions to hypothalamic activity. As you will see, a number of experimental techniques like sensory deprivation, hormone and drug administration, cortical ablation, and the production of subcortical lesions may be used fruitfully to isolate these factors. But that is only half the problem. Obviously, the factors controlling hypothalamic activity and motivation do not operate in isolation. In fact, it is quite clear that their influences interact. Therefore, it becomes an equally important problem to determine the relative contribution of each factor while the others are operating over a wide range of variation.

Experimental Evidence

Before going into the literature bearing on the operation of each of these factors in control of motivated behavior, it will help to raise a few questions that ought to be kept in mind while considering the experimental evidence. Are there different hypothalamic centers controlling each kind of motivation? Does the hypothalamus exert its influence through direct control of the final effector pathways or does it simply have a "priming" effect on effector paths controlled by other parts of the nervous system? Do all these factors operate in the control of each type of motivation or are there cases where sensory stimuli, for example, may not be important or where changes in the internal environment do not contribute? Can the same mechanism describe the control of motivation measured by simple consummatory behavior, preference, and learning? Are the same mechanisms involved in the control of simple, biological motives and complex, learned motives?

Hypothalamic centers. Review of the literature on the role of the hypothalamus in motivation brings out three general conclusions. (*a*) Damage to restricted regions of the hypothalamus leads to striking changes in certain kinds of motivated behavior. (*b*) Different parts of the hypothalamus are critical in different kinds of motivation. (*c*) There are both excitatory and inhibitory centers controlling motivation in the hypothalamus; that is, damage to the hypothalamus can sometimes lead to an increase in motivation and sometimes a marked decrease.

The evidence bearing on these three points can be summarized briefly. Many experiments have shown that restricted bilateral lesions of the hypothalamus will make tremendous changes in basic biological motivations like hunger (16, 22), sleep (49, 50, 53), and sex (6, 18, 20). Less complete evidence strongly suggests that the same kinds of hypothalamic integration is also true in the cases of thirst (61), activity (35), and emotions (5, 62). We have only suggestive evidence in the case of specific hungers (59).

It is clear that there is some kind of localization of function within the hypothalamus although it is not always possible to specify precisely the anatomical nuclei subserving these functions. The centers for hunger are in the region of the ventromedial nucleus which lies in the middle third of the ventral hypothalamus, in the tuberal region (16). (See Fig. 2.) Sleep is controlled by centers in the extreme posterior (mammillary bodies) and extreme anterior parts of the hypothalamus (49, 50). The critical region for sexual behavior is in the anterior hypothalamus, between the optic chiasm and the stalk of the pituitary gland (18, 20). The center for activity is not clearly established, but seems to be adjacent with or overlapping the centers for hunger (35). Finally, the centers for emotion are also in the vicinity of the ventromedial nucleus, perhaps somewhat posterior to the hunger centers and overlapping the posterior sleep center (50, 62).

In at least two cases it is clear that there must be both excitatory and inhibitory centers controlling motivated behavior. In the case of hunger, bilateral lesions in the ventromedial nucleus

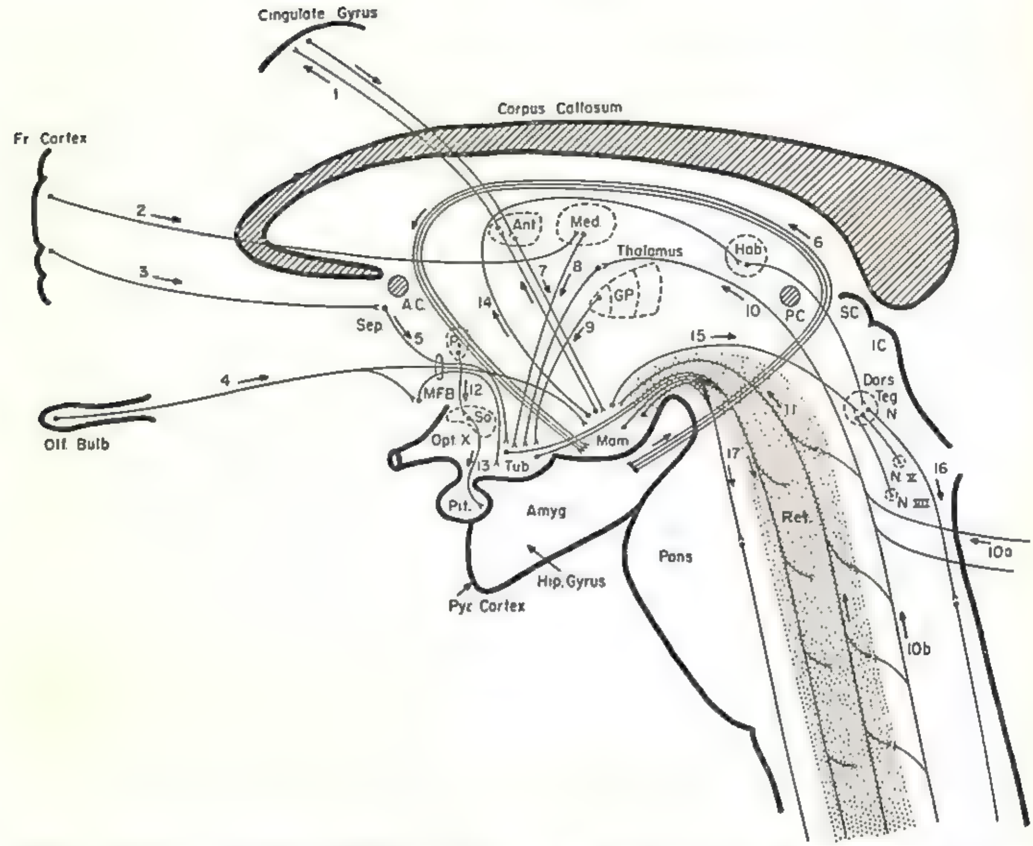

FIG. 2. Schematic drawing of the hypothalamus and its major neural connections. Adapted from W. R. Ingram's diagram in Gellhorn (30) and D. B. Lindsley's Figure 9 (43).

Abbreviations and Description of Pathways

A.C.	Anterior commissure
Amyg.	Amygdala
Ant.	Anterior thalamic nuclei
Cingulate Gyrus	Cortex of cingulate gyrus
Dors. Teg. N.	Dorsal tegmental nucleus
Fr. Cortex	Cortex of frontal lobe
GP	Globus pallidus
Hab.	Habenular nucleus of thalamus
Hip. Gyrus	Hippocampal gyrus
IC	Inferior colliculus
Mam.	Mammillary nuclei
Med.	Dorsal medial thalamic nucleus
MFB	Medial forebrain bundle
N.V	Motor nucleus, Vth nerve
N.VII	Motor nucleus, VIIth nerve
Olf. Bulb	Olfactory bulb
Opt. X	Optic chiasm
P.C.	Posterior commissure
Pit.	Pituitary gland
Pv.	Paraventricular nucleus
Pyr. Cortex	Pyriform cortex
Ret.	Reticular formation
SC	Superior colliculus
Sep.	Septal nuclei
So.	Supraoptic nucleus
Tub.	Tuber cinereum

Afferents to Hypothalamus

1. Corticothalamic fibers
2. Frontothalamic fibers
3. Frontoseptal fibers
4. Olfacto-hypothalamic tract
5. Septo-hypothalamic fibers
6. Fornix
7. Mammillothalamic tract
8. Thalamo-hypothalamic fibers
9. Pallido-hypothalamic fibers
10. Sensory systems ascending to thalamus
    10 a. cranial afferents
    10 b. somatic and visceral afferents
11. Sensory collaterals to hypothalamus
12. Paraventriculo-supraoptic fibers

Efferents from Hypothalamus

13. Supraoptic hypophyseal tract
14. Mammillohabenular tract
15. Mammillotegmental tract
16. Dorsal longitudinal fasciculus
17. Descending efferents relaying in brain stem and medulla

near the midline produce a tremendous amount of overeating (3, 16). Such a center is presumably an inhibitory one since removing it leads directly to an increase in eating behavior. On the other hand, lesions 1½ to 2 millimeters off the midline at the level of the ventromedial nucleus completely eliminate hunger behavior (3, 4). After such lesions animals never eat again, so we can call such centers excitatory centers. Supporting this interpretation is the fact, recently reported, that stimulating these lateral centers in the waking cat through implanted electrodes results in vast overeating (27). The same sort of mechanism turns up in the case of sleep. In the posterior hypothalamus, in the region of the mammillary bodies, there are excitatory centers or "waking" centers which operate to keep the organism awake (49, 50). When they are removed, the animal becomes somnolent and cannot stay awake. In the anterior hypothalamus, around the preoptic nucleus, there is an inhibitory center (49). When that is removed, the animal is constantly wakeful.

So far, only an excitatory center has been found in the case of sexual behavior. Bilateral lesions anterior to the pituitary stalk eliminate all mating behavior (18, 20), but no lesion of the hypothalamus has ever been reported that resulted in an exaggeration of sexual motivation. What little we know about the center for activity near the ventromedial nucleus suggests that it is also an excitatory center since lesions there produce only inactivity and not hyperactivity (35). In the case of emotions, the picture is not yet clear. Lesions near the ventromedial nucleus make cats highly emotional (62), and therefore this center must be inhibitory. But the lateral regions of the posterior hypothalamus seem to be excitatory, for lesions there make animals placid (50). Furthermore, direct stimulation of these posterior regions produces many of the signs of rage reactions (52).

There is some evidence that sheds light on how the excitatory and inhibitory hypothalamic centers may cooperate in the regulation of motivation. In the clear-cut cases of sleep and hunger it appears that the inhibitory centers operate mainly through their effects on the excitatory centers. At least we know that when both centers are removed simultaneously the effect is indistinguishable from what happens when only the excitatory centers are removed (3, 49). So it is convenient for present theoretical purposes to think of the inhibitory center as one of the factors which influences the level of activity of the excitatory center. In fact, to speculate one step further, it is worth suggesting that the inhibitory centers may constitute the primary neural mechanism regulating the satiation of motivation.

Sensory stimuli. What effects do sensory stimuli have upon the hypothalamus and how important are such stimuli in the control of motivation? Some answer to the first part of this question is given by the schematic outline of hypothalamic connections shown in Fig. 2. Clearly the hypothalamus has a rich supply of afferents coming directly or indirectly from all the various sense organs. In fact the diagram is really an understatement of hypothalamic connections because it is an oversimplified and conservative representation. Physiological evidence shows, for example, that there must be connections from the taste receptors via the solitary nucleus of the medulla (36). Also there is evidence of rich connections from the visual system via the lateral geniculate of the thalamus (36). There is no doubt about the fact that the hypothalamus is under very extensive sensory control.

As to the sensory control of motivation, there is excellent reason to believe that the stimuli which can set up impulses in these pathways to the hypothalamus are of particular importance. Perhaps the best example comes from the study of sexual behavior (11). The consensus of a group of studies on different mammals is as follows. Sexual behavior is not dependent upon any single sensory system. Extirpation of any one peripheral sense organ has no appreciable influence on the arousal and execution of sexual behavior. If two sensory avenues are destroyed, however, sexual behavior may be eliminated, especially in the case of the naive animal. With experienced animals, interestingly enough, it may take destruction of three sensory systems. But in neither case does it matter what combination of sensory systems is eliminated. We can conclude, therefore, that it is the sum total of relevant sensory impulses arriving at the central nervous system (hypothalamus) that is important in setting off sexual behavior.

Kleitman's analysis of sleep and wakefulness shows that the same kind of sensory control operates in this case (38). Wakefulness seems to be dependent upon the sum total of sensory impulses arriving at the waking center in the posterior hypothalamus, regardless of the particular sensory systems involved. Direct support of this kind of view is offered by Bremer's (14) physiological data which showed that maintenance of the waking rhythm of the brain is less a matter of any particular sensory input and more a matter of the amount of sensory input.

What we know about hunger and thirst suggests that the amount of motivated behavior in these cases should be a joint function of sensory impulses arising from gastric contractions or dryness of the throat and taste, tactile, and temperature receptors in the mouth. Unfortunately we have no sensory deprivation experiments that are a good test of this point. But all the evidence on the acceptability of foods and fluids of different temperatures, consistencies, and flavoring suggests the joint operation of many stimuli in the control of these types of motivation.

So far, we have mentioned only stimuli which arouse motivation. What stimulus changes could reduce motivation and perhaps lead to satiation? There are three general possibilities: (*a*) a reduction in excitatory stimuli, (*b*) interfering or distracting stimuli that elicit competing behavior, and (*c*) "inhibitory" stimuli. It is easy to find examples of the first two types of stimulus changes and to guess their mechanisms of operation in terms of the present theory. In the case of "inhibitory" stimuli, however, all we have is suggestive evidence. For example, the fact that dogs with esophageal fistulas eat (37) and drink (1, 13) amounts proportional to the severity of deprivation suggests that the stimuli which feed back from consummatory behavior might have a net inhibitory effect on motivation (see Fig. 1). Furthermore, some of the experiments on artificially loading the stomach suggest that a full gut may result in stimuli which inhibit further eating (37) or drinking (2, 13) over and above the possibility that there might be no room left in the stomach or that gastric contractions are reduced.

In summary, we can state the following working hypotheses about the sensory factors which operate in the control of motivation. (*a*) No one sensory avenue is indispensable in the arousal of motivated behavior. Instead, sensory stimuli have an additive effect on the excitability of the hypothalamus so that it is the sum total of relevant impulses arriving at the excitatory centers of the hypothalamus that determine the amount of motivated behavior. (*b*) Judging from the resistance of experienced animals to the effects of sensory deprivation in the case of sexual motivation, it seems clear that excitatory influences in the hypothalamus may be exerted by learned as well as unlearned stimuli. (*c*) There are afferent impulses to the hypothalamus which have a net inhibitory effect on the excitatory centers and thus serve to reduce motivation or produce satiation. The best guess at present is that these "inhibitory" stimuli operate by exerting an excitatory influence on the inhibitory centers of the hypothalamus. Presumably, impulses to inhibitory centers have the same kind of additive properties as impulses to the excitatory centers.

Internal environment. That the internal environment plays an important role in certain kinds of motivated behavior is a well-established fact. Two basic questions must be asked, however, before we can understand much about how the internal environment does its work. What kinds of changes that can occur in the internal environment are the important ones in motivation? How do changes in the internal environment influence the nervous system and, therefore, motivated behavior?

In terms of the present theory, we would expect the internal environment to operate in motivation by changing the excitability of hypothalamic centers. This is a reasonable expectation, for the hypothalamus is the most richly vascularized region of the central nervous system (24). Not only that, but the hypothalamus is also in direct contact with the cerebrospinal fluid in the third ventricle.

The case of sexual behavior again makes an excellent example. Experiments on the spayed, female cat (6, 17) and spayed, female guinea pig (28) have shown that hypothalamic regions must be intact and functioning if injected sex hormones are to arouse estrous behavior. If a section is made through the spinal cord only rudimentary fragments of sexual behavior can be elicited by appropriate stimulation, and injected sex hormones make no contribution to the response. Essentially the same thing is true if the section is made high in the hind brain but excludes the hypothalamus. When the decerebration is just above the hypothalamus, full estrous reactions can be aroused by appropriate stimulation, but only if sex hormones have been administered. It is clear, then, that not only is the hypothalamus the main integrating center for sexual reactions, but it is also most likely the main site of action of the sex hormones. This point is further supported by studies of female guinea pigs with pinpoint lesions of the anterior hypothalamus. These animals fail to show sexual behavior even under the influence of massive doses of sex hormones (19).

A very similar mechanism seems to be involved in the case of motivated behavior dependent upon the organism's defenses against temperature extremes (activity, nesting, hoarding, selection of high-calorie diets). We know, for example, that reactions regulating body temperature in the face of heat and cold are integrated in two separate centers in the hypothalamus (15, 51). Lesions in the anterior hypothalamus destroy the ability to lose heat and, therefore, to survive in high temperatures. Posterior hypothalamic lesions, conversely, result in a loss of heat production mechanisms so that the animal succumbs to cold. Furthermore, artificially raising the temperature of the anterior hypothalamus will quickly induce heat loss, suggesting that normally the temperature of the blood may be important in activating the hypothalamic mechanisms (15, 44). Unfortunately our information stops here. There are no direct physiological studies on the role of these temperature-regulating mechanisms in the control of motivated behavior like activity, hoarding, nesting, or food selection. But it seems clear that the temperature of the blood may be one of the kinds of changes in the internal environment that can affect the hypothalamus, and it may be important in motivated behavior.

Ample evidence demonstrates that there are important changes in the internal environment involved in other kinds of motivated behavior. In hunger it has been shown that chemicals like insulin (32, 33, 48) and d-amphetamine (57) influence the rate of eating. It is clear that these chemicals do not operate primarily through their effects on gastric contractions, but it is only by a process of elimination that we can guess that their sites of action are in the hypothalamus. Supporting this possibility is the evidence that there are chemoreceptors in the hypothalamus which are sensitive to variations in blood sugar and important in the regulation of hunger (45). In the case of specific hungers, much evidence shows that food preference and diet selection depend upon changes in the internal environment produced by such things as pregnancy, dietary deficiencies, or disturbances of endocrine glands (54). Furthermore there are some preliminary experimental data, in the case of salt and sugar appetites, to suggest that there are separate regulatory centers in the hypothalamus which are responsive to changes in salt and sugar balance (59). Finally, in the case of thirst we know that a change in osmotic pressure, resulting from cellular dehydration, is the important internal change leading to drinking behavior (31). We know further that in the hypothalamus there are nerve cells, called "osmoreceptors," which are extremely sensitive to minute changes in osmotic pressure (61). But the direct experiment has not been done to check whether or not it is these nerve cells which are mainly responsible for the control of thirst.[1]

Obviously the experimental evidence on hunger, specific hunger, and thirst is incomplete. But enough of it fits into the scheme of the theoretical mechanism proposed here to suggest the real possibility that the internal changes important in these cases operate largely through their effects on the hypothalamus.

One question still remains. What role does the internal environment play in the mechanism of satiation? About all we have to go on at present is the very striking fact from the case of specific hungers that vastly different amounts of consummatory behavior are needed to bring about satiation for different food substances. In vitamin deficiencies only a few milligrams of substance need be consumed to produce satiation, whereas in caloric deficiencies many grams of carbohydrate, fat, or protein must be ingested. Presumably, it is not the sensory feedback from consummatory behavior that is important in these cases, but rather some inhibitory effects produced by what is consumed (Fig. 1). Within the present theoretical framework, such inhibitory effects could be produced either by depression of excitatory centers of the hypothalamus or by arousal of activity in inhibitory centers. The problem is an important one and it is wide open for study.

It is clear from the foregoing that many types of motivated behavior are dependent upon changes in the internal environment. Several points are worth emphasizing. (*a*) A variety of kinds of changes in the internal environment can play a role in the regulation of motivation: variation in the concentration of certain chemicals, especially hormones, changes in osmotic pressure, and

[1] In a recent publication, Anderson of Stockholm has shown that injection of small quantities of hypertonic NaCl directly into restricted regions along the midline of the hypothalamus produces immediate and extensive drinking in water-satiated goats. (Anderson, B. The effect of injections of hypertonic NaCl-solutions into different parts of the hypothalamus of goats. *Acta Physiol. Scand.*, 1953, 28, 188–201.)

changes in blood temperature. (*b*) The best hypothesis at present is that these internal changes operate by contributing to the activity of excitatory hypothalamic centers controlling motivation. (*c*) An equally important but less well-supported hypothesis is that internal changes, normally produced by consummatory behavior, operate in the production of satiation by depressing excitatory centers or arousing inhibitory centers of the hypothalamus.

Cortical and thalamic centers. Despite the heavy emphasis laid upon the hypothalamus in this discussion, it is obvious that it is not the only neural center operating in the control of motivated behavior. In the first place, some of the sensory, motor, and associative functions of the cortex and thalamus are directly important in motivation quite apart from any influence they have on the hypothalamus. Secondly, even though the hypothalamus may be the main integrating center in motivation, it does not operate in isolation. There is much evidence that the hypothalamus is under the direct control of a number of different cortical and thalamic centers (Fig. 2).

The case of emotions offers the best example of how the cortex may operate in motivation. According to the early work of Bard and his co-workers on the production of "sham rage" by decortication, it looked as though the entire cortex might normally play an inhibitory role in emotions (5). More recent work, however, shows that cortical control of emotion is more complicated than this. Bard and Mountcastle (7), for example, have found that removal of certain parts of the old cortex (particularly amygdala and transitional cortex of the midline) produced a tremendous increase in rage reactions in cats. On the other hand, removing only new cortex resulted in extremely placid cats. Results of work with monkeys (40) and some very recent experiments with cats disagree somewhat with these findings in showing that similar old cortex removals lead to placidity rather than ferocity. The disagreement is yet to be resolved, but at least it is clear that different parts of the cortex may play different roles in the control of emotion, certain parts being inhibitory and others excitatory.

In the case of sleep, it appears so far that the cortex and thalamus play excitatory roles, perhaps having the effect of maintaining the activity of the waking center in the posterior hypothalamus. Decortication in dogs, for example, results in an inability to postpone sleep and remain awake for very long, or, as Kleitman puts it, a return to polyphasic sleep and waking rhythms (38, 39). Studies of humans, moreover, show that even restricted lesions of the cortex or thalamus alone can result in an inability to stay awake normally (25, 26). But no inhibitory effects of the cortex in sleep have yet been uncovered.

In sexual behavior it has been found that lesions of the new cortex may interfere directly with the arousal of sexual behavior (9, 11). Large lesions are much more effective than small lesions, as you might expect. Furthermore, cortical damage is much more serious in male animals than in females and is much more important in the sexual behavior of primates than it is in the case of lower mammals. On the other hand, in connection with studies of the cortex in emotions, it has been found that lesions of the amygdala and transitional cortex of the midline can lead to heightened sexuality in cats and monkeys (7, 40). So it looks as though the cortex may exert both excitatory and inhibitory influences in sexual motivation.

Evidence from other types of motivated behavior is only fragmentary, but it fits into the same general picture. In the case of hunger, it has been reported that certain lesions of the frontal lobes will lead to exaggerated eating behavior (41, 55). Hyperactivity may follow similar frontal lobe lesions and is particularly marked after damage to the orbital surface of the frontal lobe (56). The frontal areas may also be involved in what might be called pain avoidance. Clinical studies of man show that lobotomies may be used for the relief of intractable pain (29). The curious thing about these cases is that they still report the same amount of pain after operation but they say that it no longer bothers them. Presumably the frontal cortex normally plays an excitatory role in the motivation to avoid pain.

In all the cases cited so far, the anatomical and physiological evidence available suggests strongly that the main influence of the cortex and thalamus in motivation is mediated by the hypothalamus. But we do not yet have direct proof of this point and need experiments to check it.

Interaction of factors. Up to now, we have treated the various factors that can operate in the control of motivated behavior singly. However, one of the main points of the theory proposed here is that the various factors operate together in the control of motivation. Presumably this interaction of factors occurs in the hypothalamus and takes the form of the "addition" of all excitatory influences and the "subtraction" of all inhibitory influences. Some experimental evidence bears directly on this point.

In the case of sexual behavior, for example, it is clear that excitatory influences of the cortex and hormones are additive. After sexual motivation is eliminated by cortical damage it may be restored by the administration of large doses of sex hormones (10). Since the hypothalamus is the site of action of the sex hormones, it seems likely that it is also the site of interaction of the influences of the hormones and cortex.

In a similar way, it looks as though the contributions of sensory stimulation and sex hormones add in the hypothalamus. Neither hormones nor stimulation alone is sufficient to elicit sexual reactions in most mammals, but the right combination of the two will. Still another example of the addition of excitatory influences is seen in the study of the sexual behavior of the male rabbit. In this case neither destruction of the olfactory bulbs nor decortication will eliminate mating behavior, but a combination of the two operations will (21).

It is very important to know whether excitatory, and perhaps also inhibitory, influences in other kinds of motivation have the same sort of additive properties as in sexual behavior. Indirect evidence suggests they do, but direct experiments of the sort described here are needed to check the possibility.

Most encouraging in this connection is that students of instinctive behavior in inframammalian vertebrates and invertebrates have presented considerable evidence showing that sensory, chemical, and neural influences contribute jointly to the arousal of many kinds of motivated behavior (60). For example, in a number of cases it has been shown that the threshold for arousing behavior by various stimuli is lowered considerably by appropriate changes in the internal environment. In fact, in the extreme case, when internal changes are maximal, the behavior may occur in the absence of any obvious stimulation. Presumably in these cases, as in the examples of mammalian motivation, chemical and neural influences contribute to the arousal of some central response mechanism in an additive way.

The role of learning. It is obvious to every student of mammalian motivation that learning and experience may play extremely important roles in the regulation of motivated behavior. What does this mean in terms of the present physiological theory? Unfortunately, we cannot specify the mechanisms through which learning enters into the control of motivation because we are ignorant of the basic physiology of learning. But we can make some helpful inferences.

The basic hypothesis in the present theoretical framework is that learning contributes to hypothalamic activity along with influences from unlearned afferent impulses, internal changes, and cortical activity. In the case of sexual behavior we know that many animals learn to be aroused sexually by stimuli which were not previously adequate. Further, we know that in such experienced animals it is difficult to reduce sexual motivation by eliminating avenues of sensory stimulation, presumably because the extra excitatory effects produced by learned stimuli contribute to hypothalamic activity along with the impulses from unlearned stimuli. Along the same lines, it is known that sex hormones are relatively unimportant in man and in certain of the subhuman primates that have learned to be aroused by a wide variety of stimuli (12). Again, this may mean that the excitatory effects from the learned stimuli have added enough to the effects of unlearned stimuli to make it possible to dispense with the contribution of the sex hormones in arousing hypothalamic activity.

The evidence available on learning in other types of motivation fits in with this general theoretical picture, but direct physiological experiments have not yet carried us beyond the stage of inference. We know, for example, that vitamin-deficient rats can learn to show motivated behavior in response to certain flavors that have been associated with the vitamin in the past (34, 58). In fact, for a short while they will even pass up food containing the vitamin to eat vitamin-deficient food containing the flavor. Again, it looks as though flavor has become empowered by a process of learning to contribute to the excitability of the neural centers controlling motivation.

Limitations of the Theory

Like any theoretical approach, the physiological mechanism proposed here has many limitations. Fortunately none of them need be too serious as long as it is recognized that the theory is set up as a general guide for experiments and a framework for further theorizing. Obviously the theory is going to have to be changed and improved many times before it is free of limitations. In this spirit it might be said that the limitations of the theory are not much more than those aspects of motivation which need research the most. But whether we label them limitations or urgent areas of research, they deserve explicit attention.

The concept of "center." Throughout this discussion the terms "neural center" and "hypothalamic center" have been used. "Center" is a useful and convenient term, but it is also a dangerous one, for it may carry with it the implication of strict localization of function within isolated anatomical entities. Actually this implication is not intended, for it is recognized that localization is a relative matter and that no neural mechanism operates in isolation. Furthermore, it is also possible that there may be no discoverable localization of the neural mechanisms governing some types of motivated behavior. The theory simply states at the moment that the best general hypothesis is that some degree of localization of the mechanisms controlling motivation can be found in the hypothalamus.

Execution of motivated behavior. No attempt has been made in this discussion to describe the details of the efferent pathways or effector mechanisms responsible for the execution of motivated behavior. Discussion of the pathways has been omitted because we know very little about them. About all we can do at present is to guess, from anatomical and physiological studies of hypothalamic function, that the hypothalamus exerts some kind of "priming" effect on effector pathways controlled by other parts of the nervous system. Perhaps after the relationship of the hypothalamus to motivated behavior has been more firmly established we can profitably turn to the question of how the hypothalamus does its work.

A second aspect of the execution of motivated behavior has been omitted for the sake of brevity. We all recognize that an animal with certain kinds of cortical lesions, or deprived of certain sensory capacities, may be handicapped in executing motivated behavior quite aside from any effects these operations may have on the arousal of motivation. Fortunately most investigators have been aware of this problem and have taken pains to distinguish these two effects, focusing their attention mainly on the arousal of motivation. Some day, however, this theory should address the question of what neural mechanisms govern the execution of motivated behavior.

General nature of the mechanism. For theoretical purposes it has been assumed that essentially the same mechanism controls all types of motivated behavior. Obviously this is not likely to be the case, nor is it an essential assumption. In some types of motivation only parts of this mechanism may be involved, or factors not included in the present scheme may operate. For example, in some cases the hypothalamus may not be involved at all, or it may turn out that there are no inhibitory centers at work, or that internal chemical factors do not contribute significantly. There is no reason why we should not be prepared for these eventualities. But until specific experimental evidence to the contrary is forthcoming, the general mechanism proposed here still remains as the best working hypothesis for any particular type of biological motivation.

Inadequacy of behavioral measures. To a large degree the present discussion is based upon measures of consummatory behavior. We all know that the various measures of motivation are not always in good agreement, so there is good possibility that what we say about consummatory behavior may not apply to motivation measured by other methods. In fact, Miller, Bailey, and Stevenson (46) have recently shown that whereas rats with hypothalamic lesions overeat in the free-feeding situation, they do not show a high degree of motivation when required to overcome some barrier to obtain food.

Confining the present discussion mainly to consummatory behavior is clearly a weakness. But the logic behind this limited approach is to work out the physiological mechanisms in the simplest case first, and then to see how they must be revised to fit the more complicated cases.

Complex motivation. It can also be argued, of course, that the present theory is confined to the simple, biological motives. Again, it seems eminently advisable to keep the theory relatively narrow in scope until it is developed well enough to permit attack on the more complicated, learned motives.

Comparative approach. No attempt has been made here to make it explicit how the proposed theory applies to organisms representative of different phylogenetic levels. There are many obvious advantages to the comparative approach, but unfortunately, except for the case of sexual motivation, the information we have on different species is too scattered to be useful. Judging from what we have learned from the comparative study of sexual motivation, however, we can expect the various factors governing other types of motivation to contribute somewhat differently in animals at different phylogenetic levels. Certainly learning should be more important in primates than in subprimates, and the contributions of the cortex and thalamus should be greater. Much will be gained if future research in motivation follows the excellent example set in the study of sexual behavior and provides the much needed comparative data.

Advantages of the Theory

On the assumption that none of these limitations of the theory are critical, it is appropriate to ask: What is gained by proposing an explicit theory of the physiological mechanisms underlying motivated behavior? There are many positive answers to this question, and we can list some of them briefly.

Simplification of the problem. One of the main advantages of the theoretical mechanism proposed here is that it brings together, into one general framework, a number of different kinds of motivation that have been studied separately in the past. Certainly the theory encompasses the basic facts available on sex, hunger, specific hunger, thirst, sleep, and emotion. And it may also be able to handle the facts of pain avoidance, hoarding, nesting, maternal behavior, and other types of so-called instinctive behavior. As you have seen, one of the benefits deriving from this kind of simplification of the problem of motivation is the possibility of speeding up progress by applying what has been learned about physiological mechanisms from the study of one kind of motivation to the study of other kinds of motivation. Not only that, but the assumption that the hypothalamus is central in the control of all types of motivation may make it easier to explain the various types of interaction among motivations that have shown up in many studies of behavior.

Multifactor approach. Another advantage of the present theory is that it gives strong emphasis to the view that motivation is under multifactor control. Single-factor theories, so prevalent since the days of Cannon, can only lead to useless controversies over which factor is the "right" one and must always be guilty of omission in trying to account for the control of motivation. Of course, it must be stressed that the aim of the multifactor approach is not simply to list the many possible factors operating in motivation, but rather to get down to the concrete experimental task of determining the relevant factors which control motivation and the relative contribution of each.

Satiation of motivation. Unlike most previous theories of motivation, the mechanism proposed here attempts to account for the satiation of motivation as well as its arousal. In terms of the present theory satiation is determined by the reduction of activity in the main excitatory centers of the hypothalamus. More specifically, it looks as though the inhibitory centers of the hypothalamus may constitute a separate "satiation mechanism" which is the most important influence in the reduction of the activity of the excitatory centers. The possibility is an intriguing one, and it can be directly explored by experiment.

Peripheral and central control. In the past the study of motivation has been hampered by the controversy over whether behavior is centrally or peripherally controlled. The controversy is nonsense. The only meaningful experimental problem is to determine how the central and peripheral, or sensory, factors operate together in the control of behavior. It is this problem which the present theory addresses directly, and this is one of its greatest strengths.

Learned and innate control. The present theory avoids another knotty controversy by directly addressing experimental problems. Much time has been lost in psychology, and particularly in the study of motivation, in arguments over whether behavior is primarily innate or instinctive or whether it is primarily learned or acquired. The answer is obviously that it is both, and again the only meaningful experimental problem is to determine the relative contribution of each type of control. As far as the mechanism proposed here is concerned, both innate and learned factors make their contributions to the control of the same hypothalamic centers. There is still much work needed to determine the details of the mechanisms of operation, particularly of the learned factors, but some headway has been made and the problem is clearly set.

Explicit nature of the theory. Finally, a number of advantages derives simply from having an explicit statement of an up-to-date, physiological theory of motivation. In the first place, an explicit theory can serve as a convenient framework within which to organize the physiological facts we already have at our disposal. Second, the systematic organization of the facts sharply points up many of the gaps in our knowledge and suggests direct experiments that should be done in the investigation of motivated behavior. Third, an up-to-date, systematic theory provides a useful and reasonably clear conceptualization of motivation for psychologists working in other areas of research.

Summary and Conclusions

A physiological theory of motivated behavior is presented. The basic assumption in this theory is that the amount of motivated behavior is a function of the amount of activity in certain excitatory centers of the hypothalamus. The level of activity of the critical hypothalamic centers, in turn, is governed by the operation of four factors.

1. Inhibitory centers in the hypothalamus directly depress the activity of the excitatory centers and may be responsible for the production of satiation.

2. Sensory stimuli set up afferent impulses which naturally contribute to the excitability of the hypothalamus or come to do so through a process of learning.

3. Changes in the internal environment exert both excitatory and inhibitory effects on the hypothalamus.

4. Cortical and thalamic influences increase and decrease the excitability of hypothalamic centers.

Detailed experimental evidence is brought forward to show how these various factors operate in the management of different kinds of motivated behavior. The over-all scheme is shown diagrammatically in Fig. 1.

Out of consideration of this evidence a number of hypotheses are generated to fill in the gaps in experimental knowledge. All these hypotheses are experimentally testable. The ones of major importance can be given here as a summary of what the theory states and a partial list of the experiments it suggests.

1. There are different centers in the hypothalamus responsible for the control of different kinds of basic motivation.

2. In each case of motivation, there is one main excitatory center and one inhibitory center which operates to de-

press the activity of the excitatory center.

There is already much experimental evidence supporting these two general hypotheses, but it is not certain that they apply fully to all types of basic biological motivation. The hypotheses should be checked further by determining whether changes in all types of motivation can be produced by local hypothalamic lesions and whether both increases and decreases in motivation can always be produced.

3. The activity of hypothalamic centers is, in part, controlled by the excitatory effects of afferent impulses generated by internal and external stimuli.

4. Different stimuli contribute different relative amounts to hypothalamic activity but no one avenue of sensory stimulation is indispensable.

5. It is the sum total of afferent impulses arriving at the hypothalamus that determines the level of excitability and, therefore, the amount of motivation.

The neuroanatomical and neurophysiological evidence shows that the hypothalamus is richly supplied with afferents coming directly and indirectly from all the sense organs (Fig. 2). The behavioral evidence, furthermore, strongly suggests that motivation is never controlled, in mammals at least, by one sensory system, but rather is the combination of contributions of several sensory systems. Sensory control and sensory deprivation experiments are needed to check this point in the case of most kinds of biological motivation, particularly hunger, thirst, and specific hungers.

6. A variety of kinds of physical and chemical changes in the internal environment influences the excitability of hypothalamic centers and, therefore, contributes to the control of motivation.

The evidence shows that the hypothalamus is the most richly vascularized region of the central nervous system and is most directly under the influence of the cerebrospinal fluid. Furthermore, it is clear that changes in the internal environment produced by temperature of the blood, osmotic pressure, hormones, and a variety of other chemicals are important in motivation and most likely operate through their influence on the hypothalamus. Direct studies are still needed in many cases, however, to show that the particular change that is important in motivation actually does operate through the hypothalamus and vice versa.

7. The cerebral cortex and thalamus are directly important in the temporal and spatial organization of motivated behavior.

8. Different parts of the cortex and thalamus also operate selectively in the control of motivation by exerting excitatory or inhibitory influences on the hypothalamus.

Tests of these hypotheses can be carried out by total decortication, partial cortical ablations, and local thalamic lesions. It should be especially instructive to see what effects cortical and thalamic lesions have after significant changes in motivation have been produced by hypothalamic lesions.

9. Learning contributes along with other factors to the control of motivation, probably through direct influence on the hypothalamus.

10. The relative contribution of learning should increase in animals higher and higher on the phylogenetic scale.

A whole series of experiments is needed here. Particularly, there should be comparisons of naive and experienced animals to determine the relative effects of sensory deprivation, cortical and thalamic damage, and hypothalamic lesions. Presumably animals that have learned to be aroused to motivated behavior by previously inadequate stimuli should require more sensory deprivation but less cortical and thalamic damage than naive animals before motivation is significantly impaired.

11. The various factors controlling motivation combine their influences at the hypothalamus by the addition of all excitatory influences and the subtraction of all inhibitory influences.

Some experiments have already been done in the study of sexual motivation to show that motivation reduced by the elimination of one factor (cortical lesions) can be restored by increasing the contribution of other factors (hormone therapy). Many combinations of this kind of experiment should be carried out with different kinds of motivated behavior.

A number of the limitations and some of the advantages of the present theoretical approach to the physiology of motivation are discussed.

REFERENCES

1. Adolph, E. F. The internal environment and behavior. Part III. Water content. *Amer. J. Psychiat.*, 1941, **97**, 1365–1373.
2. Adolph, E. F. Thirst and its inhibition in the stomach. *Amer. J. Physiol.*, 1950, **161**, 374–386.
3. Anand, B. K., & Brobeck, J. R. Hypothalamic control of food intake in rats and cats. *Yale J. Biol. Med.*, 1951, **24**, 123–140.
4. Anand, B. K., & Brobeck, J. R. Localization of a "feeding center" in the hypothalamus of the rat. *Proc. Soc. exp. Biol. Med.*, 1951, **77**, 323–324.
5. Bard, P. Central nervous mechanisms for emotional behavior patterns in animals. *Res. Publ. Ass. nerv. ment. Dis.*, 1939, **19**, 190–218.
6. Bard, P. The hypothalamus and sexual behavior. *Res. Publ. Ass. nerv. ment. Dis.*, 1940, **20**, 551–579.
7. Bard, P., & Mountcastle, V. B. Some forebrain mechanisms involved in the expression of rage with special reference to the suppression of angry behavior. *Res. Publ. Ass. nerv. ment. Dis.*, 1947, **27**, 362–404.
8. Beach, F. A. Analysis of factors involved in the arousal, maintenance and manifestation of sexual excitement in male animals. *Psychosom. Med.*, 1942, **4**, 173–198.
9. Beach, F. A. Central nervous mechanisms involved in the reproductive behavior of vertebrates. *Psychol. Bull.*, 1942, **39**, 200–206.
10. Beach, F. A. Relative effect of androgen upon the mating behavior of male rats subjected to forebrain injury or castration. *J. exp. Zool.*, 1944, **97**, 249–295.
11. Beach, F. A. A review of physiological and psychological studies of sexual behavior in mammals. *Physiol. Rev.*, 1947, **27**, 240–307.
12. Beach, F. A. Evolutionary changes in the physiological control of mating behavior in mammals. *Psychol. Rev.*, 1947, **54**, 297–315.
13. Bellows, R. T. Time factors in water drinking in dogs. *Amer. J. Physiol.*, 1939, **125**, 87–97.
14. Bremer, F. Étude oscillographique des activités sensorielles du cortex cérébral. *C. r. Soc. Biol.*, 1937, **124**, 842–846.
15. Brobeck, J. R. Regulation of energy exchange. In J. F. Fulton (Ed.), *A textbook of physiology.* Philadelphia: Saunders, 1950. Pp. 1069–1090.
16. Brobeck, J. R., Tepperman, J., & Long, C. N. H. Experimental hypothalamic hyperphagia in the albino rat. *Yale J. Biol. Med.*, 1943, **15**, 831–853.
17. Bromiley, R. B., & Bard, P. A study of the effect of estrin on the responses to genital stimulation shown by decapitate and decerebrate female cats. *Amer. J. Physiol.*, 1940, **129**, 318–319.
18. Brookhart, J. M., & Dey, F. L. Reduction of sexual behavior in male guinea pigs by hypothalamic lesions. *Amer. J. Physiol.*, 1941, **133**, 551–554.
19. Brookhart, J. M., Dey, F. L., & Ranson, S. W. Failure of ovarian hormones to cause mating reactions in spayed guinea pigs with hypothalamic lesions. *Proc. Soc. exp. Biol. Med.*, 1940, **44**, 61–64.
20. Brookhart, J. M., Dey, F. L., & Ranson, S. W. The abolition of mating behavior by hypothalamic lesions in guinea pigs. *Endocrinology*, 1941, **28**, 561–565.
21. Brooks, C. M. The role of the cerebral cortex and of various sense organs in the excitation and execution of mating activity in the rabbit. *Amer. J. Physiol.*, 1937, **120**, 544–553.
22. Brooks, C. M. Appetite and obesity. *N. Z. med. J.*, 1947, **46**, 243–254.
23. Cannon, W. B. Hunger and thirst. In C. Murchison (Ed.), *A handbook of*

general experimental psychology. Worcester, Mass.: Clark Univer. Press, 1934. Pp. 247–263.

24. Craigie, E. H. Measurements of vascularity in some hypothalamic nuclei of the albino rat. *Res. Publ. Ass. nerv. ment. Dis.*, 1940, **20**, 310–319.

25. Davison, C., & Demuth, E. L. Disturbances in sleep mechanism: a clinicopathologic study. I. Lesions at the cortical level. *Arch. Neurol. Psychiat., Chicago*, 1945, **53**, 399–406.

26. Davison, C., & Demuth, E. L. Disturbances in sleep mechanism: a clinicopathologic study. II. Lesions at the corticodiencephalic level. *Arch. Neurol. Psychiat., Chicago*, 1945, **54**, 241–255.

27. Delgado, J. M. R., & Anand, B. K. Increase of food intake induced by electrical stimulation of the lateral hypothalamus. *Amer. J. Physiol.*, 1953, **172**, 162–168.

28. Dempsey, E. W., & Rioch, D. McK. The localization in the brain stem of the oestrous responses of the female guinea pig. *J. Neurophysiol.*, 1939, **2**, 9–18.

29. Freeman, W., & Watts, J. W. *Psychosurgery.* (2nd Ed.) Springfield, Ill.: Charles C Thomas, 1950.

30. Gellhorn, E. *Autonomic regulations.* New York: Interscience, 1943.

31. Gilman, A. The relation between blood osmotic pressure, fluid distribution and voluntary water intake. *Amer. J. Physiol.*, 1937, **120**, 323–328.

32. Grossman, M. I., Cummins, G. M., & Ivy, A. C. The effect of insulin on food intake after vagotomy and sympathectomy. *Amer. J. Physiol.*, 1947, **149**, 100–102.

33. Grossman, M. I., & Stein, I. F. Vagotomy and the hunger producing action of insulin in man. *J. appl. Physiol.*, 1948, **1**, 263–269.

34. Harris, L. J., Clay, J., Hargreaves, F. J., & Ward, A. Appetite and choice of diet. The ability of the Vitamin B deficient rat to discriminate between diets containing and lacking the vitamin. *Proc. roy. Soc.*, 1933, **113**, 161–190.

35. Hetherington, A. W., & Ranson, S. W. The spontaneous activity and food intake of rats with hypothalamic lesions. *Amer. J. Physiol.*, 1942, **136**, 609–617.

36. Ingram, W. R. Nuclear organization and chief connections of the primate hypothalamus. *Res. Publ. Ass. nerv. ment. Dis.*, 1940, **20**, 195–244.

37. Janowitz, H. D., & Grossman, M. I. Some factors affecting the food intake of normal dogs and dogs with esophagostomy and gastric fistula. *Amer. J. Physiol.*, 1949, **159**, 143–148.

38. Kleitman, N. *Sleep and wakefulness.* Chicago: Univer. of Chicago Press, 1939.

39. Kleitman, N., & Camille, N. Studies on the physiology of sleep. VI. Behavior of decorticated dogs. *Amer. J. Physiol.*, 1932, **100**, 474–480.

40. Klüver, H., & Bucy, P. C. Preliminary analysis of functions of the temporal lobes in monkeys. *Arch. Neurol. Psychiat., Chicago*, 1939, **42**, 979–1000.

41. Langworthy, O. R., & Richter, C. P. Increased spontaneous activity produced by frontal lobe lesions in cats. *Amer. J. Physiol.*, 1939, **126**, 158–161.

42. Lashley, K. S. Experimental analysis of instinctive behavior. *Psychol. Rev.*, 1938, **45**, 445–471.

43. Lindsley, D. B. Emotion. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 473–516.

44. Magoun, H. W., Harrison, F., Brobeck, J. R., & Ranson, S. W. Activation of heat loss mechanisms by local heating of the brain. *J. Neurophysiol.*, 1938, **1**, 101–114.

45. Mayer, J., Vitale, J. J., & Bates, M. W. Mechanism of the regulation of food intake. *Nature, London*, 1951, **167**, 562–563.

46. Miller, N. E., Bailey, C. J., & Stevenson, J. A. F. Decreased 'hunger' but increased food intake resulting from hypothalamic lesions. *Science*, 1950, **112**, 256–259.

47. Morgan, C. T. *Physiological psychology.* (1st Ed.) New York: McGraw-Hill, 1943.

48. Morgan, C. T., & Morgan, J. D. Studies in hunger. 1. The effects of insulin upon the rat's rate of eating. *J. genet. Psychol.*, 1940, **56**, 137–147.

49. Nauta, W. J. H. Hypothalamic regulation of sleep in rats; an experimental study. *J. Neurophysiol.*, 1946, **9**, 285–316.

50. Ranson, S. W. Somnolence caused by hypothalamic lesions in the monkey. *Arch. Neurol. Psychiat.*, 1939, **41**, 1–23.

51. Ranson, S. W. Regulation of body temperature. *Res. Publ. Ass. nerv. ment. Dis.*, 1940, **20**, 342–399.

52. Ranson, S. W., Kabat, H., & Magoun, H. W. Autonomic responses to electrical stimulation of hypothalamus, pre-

optic region and septum. *Arch. Neurol. Psychiat., Chicago,* 1935, **33**, 467–477.

53. Ranström, S. *The hypothalamus and sleep regulation.* Uppsala: Almquist and Wiksells, 1947.
54. Richter, C. P. Total self regulatory functions in animals and human beings. *Harvey Lect.,* 1942–43, **38**, 63–103.
55. Richter, C. P., & Hawkes, C. D. Increased spontaneous activity and food intake produced in rats by removal of the frontal poles of the brain. *J. Neurol. Psychiat.,* 1939, **2**, 231–242.
56. Ruch, T. C., & Shenkin, H. A. The relation of area 13 of the orbital surface of the frontal lobe to hyperactivity and hyperphagia in monkeys. *J. Neurophysiol.,* 1943, **6**, 349–360.
57. Sangster, W., Grossman, M. I., & Ivy, A. C. Effect of d-amphetamine on gastric hunger contractions and food intake in the dog. *Amer. J. Physiol.,* 1948, **153**, 259–263.
58. Scott, E. M., & Verney, E. L. Self selection of diet. VI. The nature of appetites for B vitamins. *J. Nutrit.,* 1947, **34**, 471–480.
59. Soulairac, A. La physiologie d'un comportement: L'appétit glucidique et sa régulation neuro-endocrinienne chez les rongeurs. *Bull. Biol.,* 1947, **81**, 1–160.
60. Tinbergen, N. *The study of instinct.* London: Oxford Univer. Press, 1951.
61. Verney, E. B. The antidiuretic hormone and the factors which determine its release. *Proc. roy. Soc., London,* 1947, **135**, 24–106.
62. Wheatley, M. D. The hypothalamus and affective behavior in cats. *Arch. Neurol. Psychiat.,* 1944, **52**, 296–316.

(Received February 26, 1953)

Psychological Review
Vol. 61, No. 1, 1954

THE S-R REINFORCEMENT THEORY OF EXTINCTION

HENRY GLEITMAN, JACK NACHMIAS,[1] AND ULRIC NEISSER[2]

Swarthmore College

Stimulus-response reinforcement theory as formulated by Hull (10), the most highly developed of current learning theories, has been the center of much debate and controversy. Its view of reinforcement has been challenged by the latent-learning studies (1, 32), its conception of the response has been attacked by place-learning experiments (27, 33), and its analysis of discrimination learning has been repeatedly questioned, both by the adherents of noncontinuity theories (16, 18), and more recently by other writers (28). Comparatively little attention, however, has been paid to its theory of extinction.

This omission is regrettable, in view of the fact that extinction constitutes a strategic area for any learning theory. In the first place, it represents an important phenomenon which every theory must at least attempt to explain—adaptive behavior presupposes not only the acquisition of appropriate new responses, but also the abandonment of inappropriate old ones. Furthermore, theoretical interpretations of extinction play an important part in the explanation of other phenomena; thus most S-R theorists consider discrimination learning to be the result of an interaction between excitatory and inhibitory tendencies.

This paper[3] will examine the S-R reinforcement theory of extinction, and will try to show that it suffers from some serious shortcomings. Specifically, we believe that Hull's theory of extinction does not fit all the experimental facts, involves certain conceptual difficulties, and generates some paradoxical predictions.

HULL'S THEORY OF EXTINCTION

Following Hilgard and Marquis (7), theories of extinction can be grouped into two general categories: interference and adaptation. Interference theories, such as Guthrie's (6) and Wendt's (34), assert that extinction is due to the association of interfering responses to the conditioned stimulus. Adaptation theories, such as Hull's theory of reactive inhibition, assume that extinction is caused by an inhibitory factor generated by the repeated elicitation of the response. This inhibitory factor—believed to be analogous to fatigue—is said to act against the further evocation of the response, and is usually thought to dissipate with time.

Razran (26) and Hilgard and Marquis have shown that neither of these theories by itself provides an adequate explanation of the phenomena of extinction. Interference theories fail to indicate how the interfering responses arise in the first place. They do not account for spontaneous recovery, although recent attempts in that direction have been made by Liberman (19, 20). They are further challenged by certain facts concerning the rates of conditioning and extinction. If extinction were but a manifestation of the conditioning of interfering responses, then any factor that facilitates conditioning should like-

[1] Now at Harvard University.

[2] Now at Massachusetts Institute of Technology.

[3] We wish to express our appreciation to Dr. Edward Walker for his helpful suggestions and criticisms.

wise accelerate extinction. In actual fact, stimulants increase the rate of conditioning and retard extinction while depressants retard conditioning but accelerate extinction. The negative correlation usually found between rates of conditioning and rate of extinction likewise argues against an interference theory (**7**, p. 119).

An adaptation theory alone is also inadequate. It fails to account for the fact that spontaneous recovery is usually incomplete, and that repeated extinction sessions eventually lead to a total lack of recovery. It does not explain the stimulus generalization of extinction effects nor the phenomenon of disinhibition.

Hull's theory of extinction (**10**),[4] like that presented by Miller and Dollard (**24**), utilizes both interference and adaptation concepts and thus has a considerably expanded scope. It first postulates the operation of an inhibitory factor, reactive inhibition or I_R, which tends to counteract the further occurrence of the response. This factor is assumed to result from the elicitation of the response itself, to vary with the effort involved in the performance of that response, and to decay with time. On this basis, Hull deduces a variety of phenomena such as spontaneous recovery, the superiority of distributed over massed practice, and reminiscence.

In addition to mere effector inhibition (I_R), Hull also postulates that extinction involves the production of a habit, conditioned inhibition or ${}_SI_R$—a habit of *not* responding. Its origin is explained as follows:

> . . . the after-effects of response evocation in the aggregate constitute a negative drive strongly akin to tissue injury or "pain." If this is the case, we should expect that the cessation of the "nocuous" stimulation in question or the reduction in the inhibitory substance, or both, would constitute a reinforcing state of affairs. The response process which would be most closely associated with such a reinforcing state of affairs would obviously be the cessation of the activity itself. In accordance with the "law of reinforcement" . . . this cessation of activity would be conditioned to any afferent stimulus impulse, or stimulus traces, which chanced to be present at the time the need decrement occurred. Consequently there would arise the somewhat paradoxical phenomenon of a negative habit, i.e., a habit of *not* doing something" (**10**, p. 282).

Being a habit, ${}_SI_R$ is tied to a stimulus, and presumably does not dissipate with time. It can thus be invoked to explain the generalization of extinction along stimulus dimensions, disinhibition, and the incompleteness of spontaneous recovery.

Both inhibitory factors, ${}_SI_R$ and I_R, contribute to the extinction process by summating to make up an inhibitory aggregate $\dot{I}_R$, which is subtracted from reaction potential, ${}_SE_R$, to yield effective reaction potential, ${}_S\bar{E}_R$. This relation is expressed by the following equations:

$$\dot{I}_R = {}_SI_R + I_R, \qquad {}_S\bar{E}_R = {}_SE_R - \dot{I}_R$$

It is important to note that, according to Hull, both ${}_SI_R$ and I_R are produced during rewarded as well as during unrewarded trials; the rise of the learning curve during conditioning only means that each response leads to a greater increment of reaction potential than of the inhibitory aggregate.

On the surface, Hull's theory of extinction seems to account for many of the facts with considerable elegance. Nevertheless, we believe that his conception of the extinctive process is beset by serious problems. We shall

[4] Recently, Hull's systematic formulations have been revised and elaborated in *Essentials of Behavior* (**11**), and in *A Behavior System* (**12**). Since we feel that these more recent publications have left the theory of extinction essentially unaltered, we shall base our discussion primarily upon the more familiar *Principles of Behavior*.

discuss these problems under three headings, in what we feel is the order of increasing importance: (*a*) empirical difficulties, (*b*) conceptual difficulties, and (*c*) some paradoxical predictions generated by the theory.

EMPIRICAL DIFFICULTIES

According to Hull's theory of extinction, the elimination of a response presupposes the performance of the response to be eliminated, or at least the performance of another response from which extinction effects can generalize. For both reactive and conditioned inhibition depend upon response performance, the former directly, and the latter indirectly through its dependence on I_R reduction. There are some experimental findings, however, which at least suggest that the performance of an activity is not a necessary condition for its extinction.

1. *Subzero extinction.* Evidence for such a possibility comes first from the phenomenon of "subzero extinction," demonstrated in classical conditioning. Pavlov (25) showed that when a conditioned response has been extinguished to the point of nonelicitation, further unreinforced presentations of the conditioned stimulus will nevertheless serve to strengthen extinction, as measured by a decrease in spontaneous recovery. Similar results were obtained by Brogden, Lipman, and Culler (2).

It might be argued that these effects are the result of the extinction of covert, implicit responses, which were elicited even when the overt ones were absent. Such an interpretation is consonant with the findings of Brogden, Lipman, and Culler (2) that slight forelimb movements did persist into the subzero extinction trials. This implies that crucial implicit responses or r_g's survived the elimination of the overt, "parent" responses, and that they are thus more resistant to extinction than are the latter.

2. *Latent extinction.* Further evidence comes from studies reporting an effect which might be called "latent extinction" by analogy with the phenomenon of latent learning. There have recently been three experiments in this area.

Seward and Levy (29) trained rats to run a straight alley with food on the goal platform. Subsequently, the animals were extinguished in two different ways: The experimental group was detained on the now empty *goal* platform both before and between extinction trials, whereas the control group spent equivalent periods on a *neutral* empty platform. The experimental group reached the extinction criterion in significantly fewer trials, and ran more slowly than the control group. The effect of previous detention on the empty goal platform appeared even on the very first extinction trial: compared with training there was a significant decrease in running time for the experimental animals after such treatment, whereas the corresponding decrease for the control animals was not significant. This suggests that an instrumental response can be extinguished without being elicited.

Bugelski, Coyer, and Rogers (3) took issue with the experimental design employed by Seward and Levy, pointing out that their experimental and control animals were detained on different platforms even during the extinction procedure (between trials), so that the test situation was not identical for both groups. (We don't entirely understand this objection, since Seward and Levy had already found a significant difference between the running times of their two groups on the first test trial.) Upon repeating the experiment, Bugelski, Coyer, and Rogers failed to obtain any evidence of latent extinction. There is some doubt, however, whether the repetition really duplicated the conditions of the earlier experiment, since even the control animals used by Seward and Levy gave up running sooner than did those employed in the replication. Bugelski, Coyer, and Rogers suggest that this difference may be due to an age factor; the rats used in their experiment were younger and may have been more active.

Latent extinction, however, was also obtained in an experiment by Deese (4). He trained rats to run to one side of a U maze, and afterwards was able to extinguish the correct choice response by merely placing the animals in the goal box without food. Animals who were subjected to this nonresponse extinction procedure made a smaller proportion of correct choices when again run in the maze than did control animals who were not permitted to "inspect" the empty goal box. Thus, again some extinction occurred without the prior performance of the response to be extinguished, and therefore did not seem to depend upon response-produced inhibition. In fact, nonresponse extinction was just about as effective as response extinction in producing the abandonment of the correct response. Comparing the results on four ordinary, nonreinforced response trials that were preceded in one group by four nonresponse extinction trials, and in the other group by four response extinction trials, we find hardly any difference. In other words, being *placed* in an empty goal box four times seems just as effective in reducing the likelihood of running subsequently as actually having *run* there on four occasions.

This conclusion is based upon the two groups which had eight consecutive extinction trials on the same day. The effect is even more striking with groups which were given a 24-hr. rest interval between the first four and second four extinction trials. In these groups, four exposures to the empty goal box led to a greater decrement in performance, measured on the second day, than four nonreinforced runs. In other words, response extinction showed the effects of spontaneous recovery, while "nonresponse" extinction did not.

It might be asserted that these results can be explained in a manner similar to that in which Spence (31) and others (22) have attempted to deal with latent learning; that is, by reference to fractional anticipatory goal responses—r_g's—to whose sensory consequences the turning or running responses had become conditioned during training, and which have become extinguished during the latent-extinction period. Such an explanation does not seem plausible.

Since the r_g is an implicit response, it presumably requires very little effort. It follows that many trials should be required for its extinction. As we have seen, this deduction is confirmed by the data from subzero extinction, at least if these are to be explained by the supposed extinction of implicit responses. Yet Deese's animals, given only four trials in the empty goal box, nevertheless showed extinction effects equal in magnitude to those of animals which were required to actually run to the goal box. If his results are also ascribed to the ubiquitous r_g, this now possesses somewhat contradictory properties.

The preceding discussion has not exhausted the empirical difficulties encountered by Hull's theory of extinction. To give merely two examples, the roles played by effort and by spacing in learning and extinction are by no means clear. For reasons of brevity, we have confined ourselves to the discussion of what we consider the most central empirical question: Is performance a necessary condition for the extinction of a response?

Conceptual Difficulties

The S-R reinforcement theory of extinction has shortcomings more serious than the empirical problems discussed above. These shortcomings become apparent when we try to discover how the theory's central constructs are conceptualized. We will confine ourselves here to a discussion of the *conditioned inhibition* construct, which seems to pose the most serious problems. In so doing, however, we do not wish to minimize the difficulties involved in the notion of *reactive inhibition*. The latter is usually discussed (**10**, **30**) as if it were a result of proprioceptive stimulation from the specific effectors involved in the response, yet the learned response is often defined, not in terms of specific effectors, but in the broader terms demanded by the results of place learning (**27**) and response generalization (**18**) experiments. Thus, Miller and Dollard define the response as "any activity within the individual which can become functionally connected with an antecedent event through learning" (**24**, p. 59). A related problem arises from the results of Gustafson and Irion (**21**, p. 174) and Kimble (**14**), who have shown a clear reminiscence effect in bilateral transfer. They point out that if reminiscence is to be ascribed to the dissipation of I_R, then I_R must inhibit more than a particular, specific response. We refrain from extended discussion of these matters, however, because we feel that the difficulties can probably be overcome by clarification of the definitions involved.

The concept of reactive inhibition requires more thorough consideration. Originally $_sI_R$—the "habit of not responding"—was proposed by Hull to account for the more stable aspects of extinction, and for the stimulus generalization of extinction effects. The act of not responding, from here on referred to as the "not-response," is connected to the stimulus situation by the reinforcing effects of I_R reduction. The not-response is treated as formally equivalent to an ordinary response, so that $_sI_R$ is really an S-not-R bond; thus the laws of habit formation are widened to include extinction.

In order to understand what might be meant by a "habit of not responding," we must first be clear on the meaning of the not-response itself. This consideration is all the more appropriate since the postulation of not-responses that have the same status as ordinary responses has become increasingly widespread in S-R reinforcement theory. · For instance, Dollard and Miller (**5**, p. 202), in order to subsume repression under their general theory of anxiety learning, speak of a "response of stopping thinking," reinforced by anxiety reduction.

Unfortunately, S-R theorists are somewhat vague in discussing the nature of the not-response. Sometimes Hull seems to identify it with the absence of activity, at other times with the cessation of activity, and on still other occasions he seems to assert that $_sI_R$ is I_R conditioned to the stimulus situation (all italics ours):

> Consequently there would arise the somewhat paradoxical phenomenon of a negative habit; *i.e., a habit of not doing something* (**10**, p. 282).

> Stimuli and stimulus traces closely associated with the cessation of a given activity, and in

> the presence of I_R from that response, become conditioned to this *particular non-activity* . . . (11, p. 75).
>
> The organic process most closely preceding the drive reduction would be the *cessation of the activity itself* (11, p. 75).
>
> . . . this *cessation of activity* would become conditioned . . . (10, p. 282).
>
> Stimuli closely associated with the acquisition and accumulation of inhibitory potential (I_R) become conditioned to *it* . . . (10, p. 282).

Miller and Dollard refer to a tendency to stop an activity:

> . . . Thus muscle strain and fatigue are drives constantly motivating the subject to *stop the response* he is making; escape from muscle strain and fatigue are ever present to reward *stopping*. Extinction occurs unless the effects of the drive of fatigue and consequent reward for *stopping* are overridden by the effects of other stronger drives and rewards (24, p. 40).

From the above array of quotations, no clear indication emerges as to just what it is that gets conditioned to the stimulus in $_sI_R$. However, these statements do suggest a relatively small number of alternatives. The not-response (that which gets conditioned to S in $_sI_R$) is either: (a) the *absence* of a particular activity; (b) the inhibition—*interruption or cessation*—of a response already in progress; or (c) I_R conditioned to the stimulus as a *learned drive*. We shall now examine each of these alternatives.

1. *The not-response is the absence of a particular activity.* According to this alternative, the sheer absence of a response is that which by I_R reduction will be associated with the stimulus. This conception is completely untenable. For, in this sense of not-responding, the animal is performing innumerable and indistinguishable not-responses all the time. Simultaneously with not pressing the lever in a Skinner box, he is also *not* running a maze, *not* jumping to a black card, and *not* playing three-dimensional chess. As a matter of fact, the same infinite set of not-responses is also performed when he *is* pressing the lever. Since all of these not-responses occur at the time of I_R reduction and of reward, they should all be conditioned to the stimulus. This is clearly absurd.

2. *The not-response is the inhibition—interruption or cessation—of a response already in progress.* This alternative seems to be the one most frequently implied by S-R theorists. Here it is asserted that before the not-response can be evoked, the response proper must at least have begun; that is, the animal always starts to press the lever before he stops doing so. But, in extinction he eventually fails to respond altogether, and does not even start to make the response, at least overtly. Of course, one might again suggest that when no overt response is started, there is at least an implicit one present, so that the conditions for the elicitation of the not-response as here conceived are met. Such a conception, however, raises another problem.

If implicit responses are to be utilized in S-R reinforcement theory, many phenomena suggest that these responses must be capable of being extinguished. As we have already seen, the latent-extinction results suggest such an interpretation. Furthermore, Hull's recent treatment of secondary reinforcement (11) deals with this in terms of fractional responses. Since secondary rewards can be extinguished, the r_g's must be capable of extinction. Finally, we believe that Hull's theory of problem solving (9) can be shown to require the possibility of extinguishing implicit responses. Thus, within the context of S-R reinforcement theory, r_g's must be extinguishable, and since they are conceived as formally akin to overt responses, their extinction must follow the same laws as those proposed for the latter.

How could such extinction take place? There would seem to be a need for a not-r_g to counteract the r_g. But, according to the present alternative, such a not-r_g can only be evoked after the r_g has been initiated. Once again, the complete elimination of the (implicit) response must be explained, not its interruption. Since we are already at the level of implicit responses, an even more implicit response would have to be set in motion for the purpose. This is an utterly unpalatable concept.

The present alternative, then, seems to be unsatisfactory. In order to explain the total elimination of responses, it must resort to the postulation of implicit responses. When called upon to account for the extinction of implicit responses, it becomes yet more strained.

3. *The not-response is reactive inhibition conditioned to the stimulus.* Hull (**10**) sometimes writes of $_{S}I_{R}$ as I_R which has been conditioned to a stimulus. This implies that I_R and $_{S}I_{R}$ are of the same nature, except that in the first case the inhibitory force is produced as the direct result of effector action, while in the second case the identical force is elicited by a conditioned stimulus.

Such an interpretation is formally similar to the theory of fear behavior suggested by Miller (**23**). He assumes that fear is an internal response, reflexly connected with pain, which can be conditioned to an originally neutral stimulus under suitable conditions. Kimble (**13**) has criticized the interpretation of $_{S}I_{R}$ in such terms. He argues that I_R should be treated as an intervening variable, and that responses, rather than intervening variables, become connected to stimuli.

Kimble's criticism may not be a crucial one. S-R theorists have rarely hesitated to endow their intervening variables with appropriate properties, and could perhaps treat I_R *as if* it were a response (or rather, a not-response), which can be conditioned to stimuli to generate $_{S}I_{R}$. Even this formulation, however, raises some problems. If $_{S}I_{R}$ is thought of as a conditioned I_R-response, then $_{S}I_{R}$ and I_R must have identical response properties. From this point of view, the inhibitory processes involved in $_{S}I_{R}$ and I_R must be the same in all respects save the manner in which they are aroused.

Thus, the dissipation of inhibition following the withdrawal of a conditioned inhibitor must follow the same temporal course as the dissipation of I_R, the response generalization of $_{S}I_{R}$ and I_R must be equivalent, and so on. We do not know whether S-R reinforcement theorists would be prepared to accept these consequences of the present alternative.

There remains yet a fourth possibility, that of equating the not-response with an actual activity antagonistic to the to-be-extinguished activity, i.e., making the not-response a bona fide response. The resulting theory of extinction would be rather close to the interference theories of Guthrie (**7**) and Wendt (**34**), and at the very least would have to face many of the objections that have been leveled against these. In the absence of any evidence that Hull and his coworkers had this possibility in mind, it will not be considered here.

Paradoxical Derivations from the Theory

We have tried to show that the S-R reinforcement theory of extinction encounters serious empirical problems, and contains some important conceptual difficulties. One might argue that, despite their shortcomings, the postulates of the theory permit us to deduce a great number of phenomena of extinction which actually occur. Unfortunately, however, they also necessitate certain other predictions which are clearly false.

1. *Predictions regarding the course of learning and extinction.* Hull and his co-workers believe that habit strength becomes asymptotic to a maximum value, and they usually assume that it does not decay with time. Furthermore, they assert that I_R and $_sI_R$ result as a necessary consequence of the *evocation* of the response, regardless of the presence or absence of positive reinforcement. Withholding reinforcement leads to extinction only indirectly; when no further increase in reaction potential occurs, the inhibitory action of I_R and $_sI_R$ grow unopposed.

From these assumptions it follows that the ordinary learning curve should not be monotonically increasing, but instead should rise to a maximum and then eventually return to the base line.[6] For, as the habit is repeatedly reinforced, $_sE_R$ approaches its asymptote. Once this asymptote is approximated, further reinforcements cannot add any further effective increments to the habit strength. Only I_R and $_sI_R$ can then be generated to any extent. (That $_sI_R$ is not yet at its asymptote is obvious: since extinction has not yet occurred, $_sI_R$ must be capable of further growth.) This means that from here on, further reinforcements can only lead to a *decrement* in performance, and will eventually cause the total elimination of the response. A pause between trials may at first lead to some recovery due to I_R dissipation, but this recovery will be short lived. Further trials must add to $_sI_R$ until it is approximately equal to $_sE_R$, at which point no more recovery can take place. The learning curve will have reached the base line, never to come up again. Necessarily, then, there is no learned act which can be performed for any length of time; its very repetition—regardless of reinforcement—must lead to its eventual elimination.

This prediction is at odds with everything we know about the course of learning. The phenomenon of "inhibition of reinforcement" (8) occurs only under quite special conditions, and hardly begins to do justice to this deduction. The learning curve must return to *zero*, regardless of the spacing of trials, and must do so in the *same number of trials* required for experimental extinction after $_sE_R$ has reached asymptote. One does not have to refer to experimental studies to demonstrate the fallacy of this prediction. Our daily life is full of countless activities which we perform again and again with no sign of decrement. We turn door knobs, say "how do you do," sit down on chairs, and recline on beds, and have done so since childhood. It is reason-

[6] Koch (15) notes this point in his review of Hull's *Principles of Behavior*, but does not seem to regard it as more than a matter of detail. The same problem is recognized by McGeoch and Irion (21, p. 55). They suggest that the situation could be remedied by making I subtract from N (the number of reinforced trials) rather than from $_sE_R$. In effect, this would make I_R subtract from $_sH_R$. In Hull's system, however, the indestructibility of $_sH_R$ and the merely "masking" roles of I_R and $_sI_R$ are essential; for example, they are crucial for the derivation of such phenomena as spontaneous recovery, reminiscence, and disinhibition. The suggestion made by McGeoch and Irion thus amounts to a proposal for a radically revised theory, and is not specifically relevant to the present discussion.

able to assume that such habits have reached asymptotic strength at an early age, yet there is no sign of decline.

By the same reasoning, it also follows that once a habit has been completely extinguished, reconditioning is impossible. For again, assuming that the habit strength was at asymptote prior to extinction, further reinforcement—no matter how frequent or how spaced—cannot add to it. This also is contrary to experimental fact (2) and to common observation.

The deductions just developed lead one to suspect that there is some serious flaw in the postulates which generated them. It seems to us that the problems principally derive from the assumption that there is no qualitative difference between the learning and the extinction situations, and that nonreinforcement affects performance merely by preventing the further growth of habit strength. According to the theory, an extinction trial is but a learning trial without reward—or rather with decreased reward—since I_R reduction still furnishes some reinforcement. Withdrawal of reward produces no real change in the situation. I_R and $_SI_R$ are generated during learning as well as during extinction. We shall now try to show that this conception leads to yet further paradoxes.

2. *Predictions regarding the impossibility of either learning or extinction.* In the theory of extinction originally proposed by Hull, conditioned inhibition is a habit established by reinforcement due to I_R reduction. Whenever an animal performs a response, a not-response inevitably follows it. Just how the not-response is conceived is irrelevant, so long as it inevitably occurs subsequent to the bona fide response. During the performance of the not-response, I_R dissipates. This results in need reduction, and in turn reinforces the connection between the stimulus situation and the not-response. Thus $_SI_R$ is built up. The not-response opposes the response, eventually leading to extinction.

If we accept this mechanism, we are faced with an unpleasant dilemma:

a. Extinction is impossible. Before not-responding, the animal must necessarily have responded. If I_R reduction is reinforcing, it should reinforce the response as well as the not-response. If it did, and to the same degree, extinction could not take place.

In discussing this problem, Hull (**10**, p. 301) refers to the gradient of reinforcement. He points out that the not-response is temporally more contiguous with the decrease in "nocuous" stimulation and thus to reinforcement, than is the bona fide response. In consequence, the former should be stamped in more strongly than the latter, i.e., the increment in $_SI_R$ should exceed the increment in $_SH_R$. In this manner extinction could take place. But this solution forces us onto the other horn of the dilemma.

b. Learning is impossible. As we have already seen, according to the theory, the not-response follows the response both during learning and during extinction. After pressing the lever in a Skinner box, the animal must necessarily stop pressing the lever (perform the not-response). This occurs before he reaches for the food pellet. But since the gradient of reinforcement—invoked before to make extinction possible—applies here equally, the not-response should be conditioned more strongly to the stimulus situation than should the response itself. In that case, the response can never become effectively established, since the increment in $_SI_R$ must always be greater than the increment in $_SH_R$. Any increase in the amount of reinforcement would benefit the not-response proportionately more than the response itself. Thus, learning is impossible.

Hull considers this problem also, and suggests a possible solution. He argues that the reinforcement in many experimental situations is secondary in nature, and that "this secondary reinforcement, e.g., the click of the magazine, occurs *during* the contraction and *before* the relaxation" (10, p. 302). Since reinforcement preceding a response is relatively ineffective, Hull concludes that the response would receive a greater benefit from the secondary reinforcement than would the not-response, even though the former benefits less from primary reinforcement due to I_R reduction. In this way, the not-response might receive less total reinforcement than the response proper, and the $_SH_R$ increment might outweigh the increase in $_SI_R$.

This suggestion seems inadequate, for there is no reason to assume that the response is accompanied by more secondary reinforcement than is the not-response. The effectiveness of secondary reinforcers is generally believed to be a function of their temporal proximity to primary need reduction. The not-response is necessarily closer to primary reinforcement than is the bona fide response. Hence the secondary reinforcement accompanying it should be more, rather than less. The occurrence of a consistent click at the time of the response is merely an artifact of a particular experimental condition; surely the rat would learn even if the click were made contiguous with the not-response.

We are thus left with a strange spectacle: a theory of extinction, derived from principles of learning, which must deny either the existence of learning or of extinction. The assumption of continuity between the learning and extinction situations—the failure to allow for any qualitative change brought about by withdrawal of reward—appears less and less tenable.

SUMMARY

Any theory of learning must deal with the phenomena of extinction as well as those of habit formation. S-R reinforcement theory, as presented by Hull, is one of the most influential of modern learning theories. It thus seemed appropriate to examine critically his treatment of extinction. We have tried to show that it faces a number of serious difficulties. In particular:

1. Recent experiments in the field of "latent extinction" suggest that the actual performance of a response may not be necessary for its extinction.
2. Neither reactive nor conditioned inhibition is clearly or adequately conceptualized. In particular, the "habit of not responding" has never received a satisfactory definition.
3. Certain paradoxical consequences can be derived from the theory: Not only should the learning curve inevitably decline to its starting point with continuous reinforcement, but, in fact, learning should be impossible altogether.
4. Many of these difficulties stem from Hull's assumption that withdrawal of reward introduces nothing essentially new to the situation.

REFERENCES

1. BLODGETT, H. C. The effect of the introduction of reward upon the maze performance of rats. *Univer. Calif. Publ. Psychol.*, 1929, 4, 113–134.

2. Brogden, W. J., Lipman, E. A., & Culler, E. The role of incentive in conditioning and extinction. *Amer. J. Psychol.*, 1938, **51**, 109–117.

3. Bugelski, B. R., Coyer, R. A., & Rogers, W. A. A criticism of pre-acquisition and pre-extinction of expectancies. *J. exp. Psychol.*, 1952, **44**, 27–30.

4. Deese, J. The extinction of a discrimination without performance of the choice response. *J. comp. physiol. Psychol.*, 1951, **44**, 362–366.

5. Dollard, J., & Miller, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.

6. Guthrie, E. R. *The psychology of learning.* New York: Harper, 1935.

7. Hilgard, E. R., & Marquis, D. G. *Conditioning and learning.* New York: Appleton-Century, 1940.

8. Hovland, C. I. "Inhibition of reinforcement" and phenomena of experimental extinction. *Proc. nat. Acad. Sci.*, 1936, **22**, 430–433.

9. Hull, C. L. The mechanism of the assembly of behavior segments in novel combinations suitable for problem solving. *Psychol. Rev.*, 1935, **42**, 219–245.

10. Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

11. Hull, C. L. *Essentials of behavior.* New Haven: Yale Univer. Press, 1951.

12. Hull, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

13. Kimble, G. A. Performance and reminiscence in motor learning as a function of the degree of distribution of practice. *J. exp. Psychol.*, 1949, **39**, 500–510.

14. Kimble, G. A. Transfer of work inhibition in motor learning. *J. exp. Psychol.*, 1952, **43**, 391–392.

15. Koch, S. Review of Hull's *Principles of behavior*. *Psychol. Bull.*, 1944, **41**, 269–286.

16. Krechevsky, I. A study of the continuity of the problem solving process. *Psychol. Rev.*, 1938, **45**, 107–133.

17. Lashley, K. S. Studies in cerebral functioning in learning. V. The retention of motor habits after destruction of the so-called motor areas in primates. *Arch. Neurol. Psychiat., Chicago*, 1924, **12**, 249–276.

18. Lashley, K. S., & Wade, M. The Pavlovian theory of generalization. *Psychol. Rev.*, 1946, **53**, 72–87.

19. Liberman, A. M. The effect of interpolated activity on spontaneous recovery from experimental extinction. *J. exp. Psychol.*, 1944, **34**, 282–301.

20. Liberman, A. M. The effect of differential extinction upon spontaneous recovery. *J. exp. Psychol.*, 1948, **38**, 722–733.

21. McGeoch, J. A., & Irion, A. L. *The psychology of human learning.* New York: Longmans, Green, 1952.

22. Meehl, P. E., & MacCorquodale, K. A further study of latent learning in the T-maze. *J. comp. physiol. Psychol.*, 1948, **41**, 372–396.

23. Miller, N. E. Studies of fear as an acquirable drive: I. Fear as motivation and fear-reduction as reinforcement in the learning of new responses. *J. exp. Psychol.*, 1948, **38**, 89–101.

24. Miller, N. E., & Dollard, J. *Social learning and imitation.* New Haven: Yale Univer. Press, 1941.

25. Pavlov, I. P. *Conditioned reflexes.* London: Oxford Univer. Press, 1927.

26. Razran, G. H. S. The nature of the extinctive process. *Psychol. Rev.*, 1939, **46**, 264–297.

27. Ritchie, B. F., Aeschliman, B., & Peirce, P. Studies in spatial learning: VIII. Place performance and the acquisition of place dispositions. *J. comp. physiol. Psychol.*, 1950, **43**, 73–85.

28. Saldanha, E. L., & Bitterman, M. E. Relational learning in the rat. *Amer. J. Psychol.*, 1951, **64**, 37–53.

29. Seward, J. P., & Levy, N. Sign learning as a factor in extinction. *J. exp. Psychol.*, 1949, **39**, 660–668.

30. Solomon, R. L. The influence of work on behavior. *Psychol. Bull.*, 1948, **45**, 1–40.

31. Spence, K. W. Theoretical interpretations of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 690–729.

32. Thistlethwaite, D. An experimental test of a reinforcement interpretation of latent learning. *J. comp. physiol. Psychol.*, 1951, **44**, 431–441.

33. Tolman, E. C., Ritchie, B. F., & Kalish, D. Studies in spatial learning. I. Orientation and the short-cut. *J. exp. Psychol.*, 1946, **36**, 13–23.

34. Wendt, G. R. An interpretation of inhibition of conditioned reflexes as competition between reaction systems. *Psychol. Rev.*, 1936, **43**, 258–281.

(Received December 10, 1952)

Psychological Review
Vol. 61, No. 1, 1954

PUNISHMENT: I. THE AVOIDANCE HYPOTHESIS

JAMES A. DINSMOOR

Indiana University

A possible reason for the seeming neglect of the topic of punishment in contemporary behavioral research and in most of our handbook and textbook presentations may be found in the present entanglement of theoretical treatments. So confused is the current picture that Stone, in a recent review of the literature, was led to remark that "The task of resolving apparently conflicting results . . . is an all but impossible one" (32, pp. 197–198). Actually, however, I feel that there is an available formulation which can handle the bulk of the data and which can incorporate it within a more general descriptive framework without requiring new explanatory principles. I also believe that this formulation can be shown to be consistent, at least, with those special and seemingly contradictory instances which appear to have been widely cited precisely because of the difficulties which they offer for any form of systematic treatment. I am speaking of the proposition that the main effects of punishment may be attributed to the establishment of certain avoiding reactions which prevent the completion of the original behavioral sequence.

The general suggestion that the effects of punishment may be due to some form of interfering reaction is by no means a new one. It appears as early as 1932, in some rather incidental comments by Thorndike. At that time, Thorndike presented a summary of several studies which seemed to indicate that punishment had little or no effect on the preceding behavior. However, he recognized the necessity of providing some kind of an "escape clause" to deal with those special, as it seemed to him, and anomalous cases where the punishment of one response *did* facilitate the elimination of this response and the acquisition of a nonpunished alternative. To deal with such observations he offered the suggestion that this effect was due, not to a direct weakening of the punished response itself—as he had previously postulated in the law of effect—but to the strengthening of the alternative reaction. "The person or animal is led by the annoying aftereffect to do something else to the situation" (33, p. 311). Or later, "The idea of making [the] response or the impulse to make it then tends to arouse a memory of the punishment and fear, repulsion, or shame. This is relieved by making no response to the situation . . . or by making a response that is or seems opposite to the original response" (34, p. 80).

Stemming from Thorndike's approach, we have such later developments as Guthrie's contiguity interpretation (e.g., 11), Estes' "anxiety" state (9), and various references to "heightened tension" (13, pp. 245–246) or to inferred drives of "fear" or "anxiety" (e.g., 8, 16) which are said to be reduced by making an opposing or conflicting response. In particular, several authors have at least mentioned an avoidance interpretation, the fullest treatments being those by Dollard and Miller (8, pp. 75–76), Mowrer (16, pp. 91, 118, 154, 210, 262 ff.; 17), Mowrer and Kluckhohn (18, pp. 80–81), and Skinner (31, esp. pp. 188–189); but even these are obviously rather brief. Furthermore, no attempt has yet been made to lay a detailed and comprehensive statement of the hypothesis alongside the published findings from empirical studies of punishment to see how well such a statement fits the known facts.

In this paper I will merely outline the hypothesis itself. First, I will review some of the empirical studies of secondary aversive stimulation and avoidance training in order to see what principles are required for their interpretation. Next, I will compare the experimental operations used in avoidance training with those used in a study of punishment, in a free responding situation. Finally, I will try to show what should or must happen when we apply an aversive stimulus following successive instances of a given response.

Aversive Stimuli

Since the concept of an aversive stimulus is fundamental to subsequent discussion, I will begin by offering a definition. I have selected the word *aversive* both for the frequency of its appearance in the experimental literature and for the strength of its behavioral connotations in everyday usage. (In Webster's International Dictionary [2nd Ed.], for example, *aversion* is first defined as "act of turning away" and *avert* is further defined as "to cause to turn away" or "to ward off, or prevent, the occurrence or effects of.") I will use the word in a strictly functional or behavioral sense, with no reference to its subjective properties or to any assumed drive which might be said to be aroused or reduced by the presentation or removal, respectively, of the stimulus. It will refer to a class of stimuli which are suitable for studies of "escape training" (13) or "aversion" (14). The critical observation is that *the reduction or elimination of the stimulus increases the frequency or probability of the preceding behavioral sequence*—that is, that it is reinforcing to the subject.

For the naive organism, this classification apparently includes such stimulating events as immersion in water and certain intensities of light, sound, temperature, and electric shock. This is not to say, however, that an aversive stimulus cannot be stripped of its original properties or that these cannot be overlaid by other properties acquired through special training or instruction. In practice, most of the empirical studies on avoidance and punishment have been based on the administration of shock to rats, although occasional reference will be necessary to other stimuli or other organisms.

How Neutral Stimuli Become Aversive

What happens when a neutral or ineffective stimulus is paired with one which is already aversive to the subject, such as shock? A relatively clear and simple answer to this question may be found in an experiment by Brown and Jacobs (2, Experiment II). The apparatus consisted of two adjoining compartments, each with a shock grid as a floor, which were separated by a two-inch barrier surmounted by a guillotine-type door. In the first stage of the experiment, the experimental animals (rats) were each given ten presentations of a pulsating light and tone paired with a pulsating shock. Each presentation consisted of nine seconds of light and tone, overlapping with a final six seconds of shock. No systematic means of escape was provided during this stage of the experiment.

The second step was to test the functional properties which had been acquired by the light and the tone as a result of their pairing with shock. Forty trials were given. On each trial the door was raised and the light and tone were presented without further shock. When the animal passed over the hurdle from one compartment to the other the light and tone were turned off and the door was lowered behind him. The time required for the animal to respond was measured on successive trials.

A group of control animals, which had not been shocked, were found to run somewhat less promptly from trial to trial. But the experimental animals ran more and more quickly; their latencies showed a rather sharp drop for the first 16 or 20 trials. Later, however, a slight, but significant, rise appeared. The early decline in latency shows that the removal of the light and tone was a reinforcing operation which strengthened the response of running from one compartment to the other. The final rise in latency presumably reflects the gradual loss which occurs in the effectiveness of secondary aversive stimulation when it is no longer paired with the primary stimulus.

An attempt has been made by Barlow (1) to specify more exactly what is the necessary temporal relation between the primary and secondary stimuli, and a related study has been conducted in avoidance training by Mowrer and Suter (16, pp. 280 ff.). These studies both suggest that the critical relationship is between some phase of the secondary stimulus and, more specifically, the beginning or onset of the primary stimulus, such as shock. Similarly, the effects of presenting the secondary stimulus without the accompanying shock have been isolated and separately investigated in experiments by Schoenfeld and Antonitis (25) and by Page and Hall (23). These studies indicate that such a stimulus loses its aversive character when it is no longer paired with the primary stimulus.

On the basis of several replications, then, the main fact seems to be reliably established: that a neutral stimulus which is presented just prior to or overlapping with the administration of a primary aversive stimulus, like shock, acquires an aversive property in its own right and becomes what we may call a conditioned or secondary aversive stimulus. When we try to make use of this stimulus as a reinforcing agent, however, a difficulty arises. The reinforcing operation—terminating the stimulus without shock—is incompatible with the establishing and maintaining operation, pairing the stimulus with shock. When it is terminated, therefore, without being paired with the primary stimulus, our secondary stimulus gradually loses its effectiveness. The temporary nature of the secondary aversive property might seem to limit the role which these stimuli can play in the maintenance of behavior over an extended period of time. The difficulty is readily resolved, however, if the pairing is restored whenever the stimulus is weakened and the animal fails to respond within an arbitrary time limit. This is the basic paradigm for what is known as avoidance training.

How Avoiding Reactions Are Maintained

In studies like those we have just been considering, two separate and distinct operations have been employed in successive phases of the experimental procedure: (*a*) the secondary stimulus is paired with the primary stimulus, and (*b*) the termination of the secondary stimulus is used to reinforce a selected response. In a simple and relatively effective form of avoidance training, these two procedures are interspersed or interwoven. At the beginning of each trial a secondary stimulus or "warning signal" is presented by the experimenter. If the animal makes the required response the signal is terminated; but when the animal fails to respond within an arbitrary time limit, the primary stimulus is applied.

As an example of this form of training, let us take "Group III-On-Run" from an experiment by Mowrer and Lamoreaux (16, pp. 126 ff.; 20). The warning signal was a change in the pattern of illumination produced by turning on two overhead lamps and turning off a single lamp beneath the grid. Five seconds of grace were allowed in which the rat could run to the opposite end of the alley or shuttle box. If he did this, the signal was terminated, or changed back, and no shock was applied; if he did not, the stimulus was followed by two seconds of shock. On the first day of training, the two animals in this particular subgroup made the response only twice apiece in 10 trials; but on the eleventh and twelfth days, both animals ran on all 10 trials. Thus, these animals learned to run to the opposite end of the apparatus when the only direct reinforcement for this act was provided by the change from a pattern of stimulation that was otherwise followed by shock. The results were similar for other subgroups. Although the procedure which is used in avoidance training is relatively complicated, the general effects of this procedure can be predicted from a study of the way in which neutral stimuli are made aversive or stripped of their aversive character.

If the conditioning of avoiding responses is based on the termination of secondary stimuli, and if the effectiveness of these secondary stimuli is based on their pairing with the shock, it follows that some limit must be set on the frequency with which the avoiding response will be made. When the animal makes the required response for several trials in succession, the pairing operation is interrupted and the effects of successive stimulus-terminations should dwindle. That this is indeed the case is suggested by more detailed observations of the subjects' behavior under fairly comparable conditions reported by Sheffield (26). Here the warning signal was a two-second tone (apparently of fixed duration), with the shock coming in the last tenth of a second. Guinea pigs were used as subjects. If the animal turned a rotating cage or activity wheel by 1 in. before the shock was due, the shock was omitted. "With successive omissions of the shock," Sheffield reports, "the amplitude of the conditioned response tended to decrease and latency to increase, until the animal failed to turn the wheel the required inch in the required time, and another shock was received. As training continued, more and more successive conditioned responses occurred without requiring [shock], but extinction between [shocks] continued throughout the training" (26, p. 171). The amplitude was restored, of course, and the latency cut down once more after a trial on which the animal failed to respond and the shock was actually administered.

AVOIDANCE TRAINING WITHOUT A SIGNAL

Actually, no independent warning signal need be presented by the experimenter. This is demonstrated by some data recently reported by Sidman (27, 28). In this study, one-fifth of a second shocks were administered to the rat at regular intervals of time ("shock-shock interval"). The animal was permitted to avoid or delay the shock, however, by pressing a bar or lever at one end of the chamber. When he did so, the shock was postponed for another interval—sometimes the same, sometimes different—which was timed from the beginning of each successive response ("response-shock interval"). Some fifty animals were successfully conditioned by this procedure.

In an extension of his preliminary work, Sidman conducted three of his original animals through an extended series of training sessions at a variety of response-shock and shock-shock intervals. Regular changes in the rate of responding were obtained from each of these animals as a function of the length of either interval. In general, the more frequently the animals were shocked, the more frequently they responded. The rate of responding rose with shorter and shorter shock-shock intervals and, up to a point, with shorter and shorter response-shock intervals. At relatively short response-shock intervals, however, a new phenomenon appeared: the original function gave way, and a "delay-of-punishment" gradient took over. That is, the rate began to decline with very short intervals between the response and the subsequent shock, and the response tended to disappear when the shock followed almost immediately.

AN INTERPRETATION OF AVOIDANCE TRAINING

In interpreting Sidman's data (and later, the operation of punishment) we are forced, in order to construct a general or inclusive description, to make an appeal to stimuli which are: (*a*) not clearly specified; (*b*) not readily observed or recorded; and (*c*) most important, not under the direct control of the experimenter. We might, by analogy, call the stimuli which may be presented or withheld at the will of the experimenter "independent" stimuli and those which are produced by the subject without the intervention of the experimenter "dependent" or "intervening" stimuli, all in accord with the application of these adjectives to the noun "variable." If the reader objects to an appeal to such stimuli, there are two ways in which he may place them, in a sense, under the control of the experimenter. First, he may invade the organism by surgical or pharmacological techniques and presumably segregate or insulate the subject, in a manner of speaking, from the consequences of his own behavior. Or, as an alternative, he may add or subtract presumably equivalent stimuli to or from his experimental operations, making them likewise contingent upon the subject's response. Then we can see what effect these substitute dependent stimuli have on the over-all pattern of behavior. This procedure might well be used to substantiate Sidman's interpretation of his results.

Sidman's interpretation may be paraphrased as follows. Any form of behavior other than pressing the bar will eventually be followed by shock. The dependent stimuli that accompany such behavior thereby acquire an aversive character, through their pairing with the primary stimulus. But pressing the bar is never immediately followed by shock if a reasonably long response-shock interval is used, and the stimulation which accompanies this form of response does not become aversive. Hence, whenever a bar press follows some response that has previously been shocked, it will be reinforced by the change from an aversive to a nonaversive pattern of stimulation. At first only a few of the possible forms of response may have been shocked, while other responses have not been so paired. At this stage, the bar pressing is not always reinforced, and the rate of pressing fluctuates. As the training continues, however, more and more of the animal's behavioral repertoire is shocked, with the result that the avoiding response is fairly regularly reinforced. A ceiling is finally imposed by the very success of the avoiding response itself, as in other studies, which reduces the over-all frequency of shock and thereby limits the frequency of pairing and the effectiveness of a given change in stimulation.

We now see that the warning signal, which is usually provided in a study of avoidance training, plays a relatively subtle role. There are three possible relationships which tend to be confounded and which are very difficult to segregate in a given study:

1. As in Sidman's procedure, the stimuli produced by *S* himself in making responses other than that prescribed by *E* are to some extent correlated with the shock and presumably provide some basis for the reinforcement of the avoiding reaction. This may be *one* of the sources for the responding that occurs in the absence of the signal, between successive trials (e.g., 19, 21). But these stimuli are in actuality paired with the shock only when they are accompanied by the signal. Thus, the "true" or most effective secondary stimuli are a set of stimulus combinations or compounds (24), each including the warning signal as one of its elements. The effectiveness of these compounds presumably depends on the exact temporal relation between the signal and the shock (e.g., 16, pp. 280 ff.; 37).

2. Furthermore, the effectiveness of the training procedure seems to depend also on whether both elements of these compounds are simultaneously terminated by the avoiding response; if the signal element is terminated before or after the response, the change in stimulation produced by the response is somewhat smaller and less discriminable, and the reinforcement is less effective (16, pp. 84 ff.; 19, 21). In avoidance training, as such, the relation between the signal termination and the *response* is necessarily confounded with the temporal relation between the signal and the *shock;* but it may be separated in a study of secondary aversive stimulation, where the operations of stimulus pairing and reinforcement have been segregated.

3. Finally, the warning signal may "set the occasion" for the (maximal) reinforcement of the avoiding response. That is, it seems to act like a cue (8) or discriminative stimulus (14, 29, 30). Early in training the rate of responding may be about the same in the presence of the signal or in its absence (between "trials"); but as the training continues, more responses occur in the presence of the signal and fewer occur in its absence, *provided that an opportunity is given for the animal to make these nonreinforced responses* (3)—*i.e., for extinction of responses in the absence of the signal.*

Mowrer and Lamoreaux have discussed this problem in somewhat similar terms, and conclude that it is the necessity for forming a discrimination of some type between the presence and absence of the signal which slows down the acquisition of avoiding responses under the customary procedures (21).

A Comparison of Procedures

Now that we have seen how the subject learns to prevent the arrival of the shock in a study of avoidance training, we are in a position to apply these principles to the inhibition or suppression of the response which is observed in most studies of punishment. Actually, the two procedures are so similar that it is difficult to find a justification for any major distinction in theoretical treatment. As Mowrer says, such a distinction is "far from parsimonious, not to say an outright contradiction" (17, p. 421). What distinctions there are, of course, arise from the fact that in avoidance training the experimenter selects the response that shall be required to *avoid* the shock; whereas in an experiment on punishment he specifies and records the response that *produces* the shock.

This does, however, lead to certain consequences which may be worth inspecting. First, in avoidance training the class of responses which are *not* followed by shock is extremely narrow; it includes but one form—that specified by the experimenter as the avoiding response. But in a free responding situation the class of responses which *will* eventually be paired with the shock is extremely broad, including anything else the animal might do. The breadth of definition for these two classes of response is also reflected in the initial frequencies of behavior: before conditioning, the frequency of the avoiding response is likely to be relatively low, to constitute a small part, quantitatively as well as qualitatively, of the animal's activity; the combined frequency of other forms of behavior will be relatively high.

The situation is reversed in a study of punishment. Here, it is the class of responses which *are* followed by shock, for example, which is limited to a single behavioral sequence or chain. And in the usual experimental situation the initial frequency of this sequence is very low, so low, in fact, that it is ordinarily necessary to provide some form of reinforcement to boost its rate to a level where the inhibition may readily be observed. But the class of responses which are *not* followed by shock is relatively broad, including any form of response which conflicts with members of the punished sequence; and these responses are already quite plentiful at the beginning of training.

Second, the experimenter has chosen different criteria for the administration of the shock in the two cases, and this alters the detailed response-shock contingencies both for the avoiding responses and for the punished responses. In avoidance proper, he delivers the shock at regular intervals whenever the animal *fails* to make the required response. He does not specify the exact relationship between other forms of response and the arrival of the punishment. A given alternative, therefore, need not immediately or invariably be accompanied by shock unless this is continuous, as in simple escape training. Thus, a certain amount of time will be required before each of these responses has effectively been paired with the punishment.

But in a study of punishment itself, the shock is directly contingent upon *making* a particular response. The pairing can be made immediate and invariable, unless the experimenter himself wills it otherwise. Special schedules, such as delayed or intermittent punishment, may readily be imposed.

Similarly, in a study of avoidance training the experimenter not only decides that a certain form of response shall lead to avoidance of the shock, he also determines *how long* the shock will be postponed following this response, if it is not repeated—as illustrated by Sidman's "response-shock interval" or by the customary interval between trials in a signal study.

Not so with punishment: Here, the relationship between the avoiding responses and the shock is less direct, for it depends on what these do to the original sequence of behavior. The consequent variation in the delay of punishment has a selective effect on various forms and durations of avoiding response (27, 28, 36, 37). The way is opened for a "shaping up" or differentiation of the original avoiding behavior.

Finally, the observations are different. In an avoidance study the experimenter has defined, by his criterion for administering or withholding the shock, the form of the avoiding response. This turns out to be the only response which can readily be recorded, since the alternative forms have not been defined and a specified alternative may not be inclusive. In a punishment study, on the other hand, it is the punished response which has been defined, and the avoiding responses cannot readily be recorded. In either experiment something might be gained by recording one response as a representative of the class of behavior which is not ordinarily observed, although the frequency of any single response is likely to be too low, unless experimentally reinforced, to provide a very sensitive index to the remainder of the animal's behavior.

DISCRIMINATION OF THE AVOIDING REACTIONS

There is one special problem in applying avoidance theory to the action of punishment which has not been explicitly discussed by previous writers. The behavior which is punished constitutes only a small fraction, qualitatively speaking, of the animal's total repertoire. In a laboratory study, to be sure, this behavior may have been strengthened to such an extent by direct experimental reinforcement that it constitutes a relatively large proportion, *quantitatively* speaking, of the animal's activity. In this case, elements of the punished sequence will intrude at frequent intervals between the avoiding responses. Most of the animal's behavior should remain "relevant" to the punished sequence. If it is the pressing of a bar which is punished, for example, he should spend the bulk of his time in the vicinity of this bar. He will not "get very far," to judge from analogous data (15), from the punished act. Since under these circumstances the animal is almost always in danger of being punished, no special timing or discrimination of his avoiding responses may be required.

But in this respect the laboratory does not necessarily mirror life. Outside of the laboratory a given sequence may not have such a degree of strength. The animal may spend much of his time in activity which is essentially "irrelevant" to the punished response, i.e., which shows no major change in frequency following the institution of punishment. The punished responses intrude only occasionally among a variety of other forms of behavior. If we assume that there is some over-all limit to the frequency of the avoiding responses, a discrimination seems to be necessary. For if discrete avoiding responses were interspersed at random, regardless of what the animal might be doing, they would not conflict temporally with the appearance of the response which is punished and should have relatively little effect on its frequency.

In order, then, to inhibit or suppress the punished response, the avoiding responses must in a sense *anticipate or forestall it* by arising at just the moment when this response itself would otherwise appear. They must, we might say, be correlated with its expected occurrence. No "expectancy" construct, however, is required. The problem is much the same as the problem of accounting for the proper timing of avoiding responses to a signal, so that they may appear just prior to the primary stimulus. And the answer to this problem, too, is quite analogous.

CHAINING

Neither our own everyday behavior nor the activity of one of our subjects in the laboratory is made up, in atomistic fashion, of a random series of discrete and unrelated acts. Experimentally reinforced behavior, in particular, flows along in fairly orderly and regular sequences or "chains" (14, 30), as may be established by the most casual observation. Most of our laboratory records, it is true, depend on timing or tallying a single response, such as pressing a bar, turning to the right, or entering a goal box. We do not and cannot record and quantify everything that the animal does. This should not lead us, however, to ignore, where relevant, the fact that the behavior which we are studying in the "modified Skinner box," the T maze, or the runway actually consists of a continuous flow of activity from which we have rather arbitrarily abstracted a single, readily recorded element.

Again we are forced to consider stimuli which are not directly under the control of the experimenter, for each of the actions in a behavioral sequence has some effect on the current stimulation. An action may enlarge, contract, add, subtract, or otherwise alter some set of visual stimuli as the animal turns his head or moves about; it may bring him into physical contact with some object in his environment, such as a bar, a pellet, a barrier, or a wall; it may produce apparatus noises or bring new odors; or, as a minimum, it will normally produce a certain amount of proprioceptive stimulation. Although these stimuli arise in the chain as a natural consequence of the animal's own behavior, without any special intervention by the experimenter, we can largely duplicate their relationships to a particular response or their own interrelationships by direct experimental manipulation. Work of this sort has been conducted largely under the headings of discrimination training and secondary reinforcement.

DISCRIMINATIVE AND REINFORCING STIMULI

A given chain is completed and reinforced only when the necessary members occur in the proper sequence or order. It will not do, for example, for the animal to go through the motions of pressing a bar when it is at the opposite end of the cage, or to chew before the pellet is in the mouth. The function of signalling, so to speak, when to make a given response, or of "setting the occasion for" this response, is performed by the stimulus elements in the chain. There is a three-term relationship here: discriminative stimulus— response → reinforcement. It is only in the presence of the discriminative stimulus, as Skinner has called it (30, 31), or S^D, that the next response in the chain is appropriate and actually leads to a reinforcing state of affairs. This relationship is probably well known to most of my readers and need not be labored here. Empirical demonstrations are numerous. They show that when the reinforcement of a given response—e.g., pressing a bar—is made to depend on the prior presence of a certain stimulus, be it wholly arbitrary, the animal comes to make the response quite promptly (4, 29, 30), or with increased frequency (5), when the stimulus appears, but fails to respond with any great frequency when this stimulus is absent.

One of the stimulus functions, then, which is crucial to the formation of a chain, is the acquisition by this stimulus of discriminative properties. In addition it would appear that such stimuli also acquire reinforcing properties, along with their discriminative role, so that they also serve to maintain the strength of the response which produces them (e.g., 10, 30). Although the acquisition and loss of this property seem to be governed by the same factors which govern the acquisition and loss of a discriminative property (4, 7, 22, 39), we customarily refer to these stimuli—while exercising their reinforcing function—as secondary reinforcers.

Let us now consider what happens when an aversive stimulus like shock is applied as a punishment following some particular member of the chain. Again we have a three-term relationship: discriminative stimulus—response → aversive stimulation. The punished response follows upon its appropriate stimulus; the punishment itself follows upon the response; thus, *through the mediation of the animal's own behavior, aversive stimulation is paired or correlated rather specifically with the discriminative stimulus for the punished response.* If, furthermore, the entire chain is run off fairly regularly and fairly swiftly, the aversive stimulation may also follow rather closely upon some of the stimuli which appear earlier in the sequence. And finally, it is closely associated with whatever stimulation may arise during the execution of the punished act itself (12, 16, p. 262; 17). This reduces to the same analysis if we break down what we had hitherto regarded as a single act into a more detailed sequence or chain in its own right.

These stimuli, then, play a role which is similar to that of the "warning signal" in the conventional study of avoidance training. Patterns of stimulation which include these elements are more closely and more frequently paired with the shock, and should be more effective as aversive compounds; responses which terminate these elements should be maximally reinforced; and by "setting the occasion" for maximal reinforcement, these stimulus elements should serve as cues or discriminative stimuli for the avoiding responses. (We do, in fact, find that arbitrary stimuli which indicate the punishment or nonpunishment of a given response *do* affect its frequency [6].) In a sense, then, these are not only discriminative and reinforcing stimuli for members of the chain but *discriminative and reinforcing (i.e., by their termination) stimuli for a corresponding set of avoiding reactions.*

DIFFERENTIATION OF THE AVOIDING REACTIONS

Just as the animal must learn to make his avoiding responses at the *time* when the punished response is about to occur, to make them temporally incompatible, so he must also learn to make his responses of such a *form* that they are physically or topographically incompatible with the punished response (or with earlier members of the chain). It is obvious how this occurs. The pairing between the discriminative stimuli and the punishment is mediated, as I have said, by the animal's own behavior on continuing with the chain and making the punished response. The avoiding responses are reinforced precisely *because* they are incompatible with the original sequence; otherwise, they too would be followed by shock. While we cannot specify the exact form which these new responses will take, we can, from our knowledge of the basis of their reinforcement, make some tentative predictions.

First, the animal may halt, "freeze," or hold a pose. This probably involves a fine-grain vacillation between incipient movements toward completing the chain and opposing movements which serve to restore the original position (38). There may be some tendency for the animal to hold these positions for longer and longer durations (12, Experiment 3), as this further delays the punishment. But this development will be limited by the strength of the original chain (35, 38), and the mean duration of these holding responses will presumably reflect most of the variables which influence the original rate of the punished response.

Again, the animal may make responses which are incompatible with the next member of the chain and serve as digressions from the sequence. Even a slight delay in the completion of the chain may to some extent be reinforced, and a certain amount of seemingly pointless "boondoggling" may be expected, like the dilatory behavior of a small child heading for bed. The animal may scratch himself, stand on his hind legs, "wash his face," or push the sawdust about. If these responses do nothing to cancel the previous member of the chain, however, they may well be followed by immediate completion of the original sequence.

A certain premium is therefore placed on those forms of response which "undo" or cancel out one of the members of the sequence by an opposing movement or a reversal of the progression. The animal may let the bar come up again; he may drop or let go of some object; he may turn his head away from the visual stimuli; or he may withdraw bodily from the locus of the punishment. These responses remove the most important elements of the aversive pattern, namely, the discriminative stimuli for the next response in the chain. Furthermore, they "set him back" in his progress, so that he is forced to repeat one or more members of the chain to get back to the point where he was before. Thus, some differentiation of the *form* of the avoiding responses would seem likely, on the basis of selective reinforcement—variations in the temporal interval between the response and the punishment (27, 28, 36, 37). Lengthy sequences of incompatible responding, such as wandering to the opposite end of the cage, might be strengthened to some extent if the original sequence is weak; but these too are limited by the tendency to return to the chain. If the punished behavior is relatively strong, they may even be "crowded out" by the combined interference resulting from the original chain plus more localized avoiding responses. The over-all situation is reminiscent of the "equilibrium" studied by Miller, Brown, and Lipofsky (in 15), although their analysis is limited to a somewhat specialized situation.

SUMMARY

By punishing an animal for making a given response (that is, by applying aversive stimulation), we can reduce its frequency of occurrence. The purpose of this paper has been to show how we can fit this observation into a more general theoretical framework without adding new and independent principles to our system. Accordingly, I have tried to deduce the main effects of punishment from the principles already demonstrated in studies of secondary aversive stimulation and avoidance training. In general, my hypothesis has run as follows: The punished response is not an isolated incident, *in vacuo*, but a member of some sequence or chain of responses which is linked together by a series of discriminative, and thereby secondary reinforcing, stimuli. The stimuli which come immediately before the punished response are paired by the response itself with the ensuing punishment. By virtue of this pairing, they gain an aversive property in their own right. Any form of behavior which is incompatible with some member of the chain and delays the completion of the sequence will be reinforced, and thereby conditioned and maintained, by the corresponding elimination or transformation of these conditioned or secondary aversive stimuli. These responses are functionally equivalent to the responses which are investigated in a formal study of avoidance conditioning. The fruitfulness of this hypothesis may therefore be tested by a detailed comparison of the functional relations observed in studies of punishment and studies of avoidance training.

REFERENCES

1. Barlow, J. A. Secondary motivation through classical conditioning: one trial nonmotor learning in the white rat. *Amer. Psychologist*, 1952, **7**, 273. (Abstract)
2. Brown, J. S., & Jacobs, A. The role of fear in the motivation and acquisition of responses. *J. exp. Psychol.*, 1949, **39**, 747–759.
3. Coppock, H., & Mowrer, O. H. Intertrial responses as "rehearsal": a study of "overt thinking" in animals. *Amer. J. Psychol.*, 1947, **60**, 608–616.
4. Dinsmoor, J. A. A quantitative comparison of the discriminative and reinforcing functions of a stimulus. *J. exp. Psychol.*, 1950, **40**, 458–472.
5. Dinsmoor, J. A. The effect of periodic reinforcement of bar-pressing in the presence of a discriminative stimulus. *J. comp. physiol. Psychol.*, 1951, **44**, 354–361.
6. Dinsmoor, J. A. A discrimination based on punishment. *Quart. J. exp. Psychol.*, 1952, **4**, 27–45.
7. Dinsmoor, J. A. Resistance to extinction following periodic reinforcement in the presence of a discriminative stimulus. *J. comp. physiol. Psychol.*, 1952, **45**, 31–35.
8. Dollard, J., & Miller, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.
9. Estes, W. K. An experimental study of punishment. *Psychol. Monogr.*, 1944, **57**, No. 3 (Whole No. 263).
10. Ferster, C. B. Sustained behavior under delayed reinforcement. *J. exp. Psychol.*, 1953, **45**, 218–224.
11. Guthrie, E. R. *The psychology of learning.* (Rev. Ed.) New York: Harper, 1952.
12. Hefferline, R. F. An experimental study of avoidance. *Genet. Psychol. Monogr.*, 1950, **42**, 231–334.
13. Hilgard, E. R., & Marquis, D. G. *Conditioning and learning.* New York: Appleton-Century, 1940.
14. Keller, F. S., & Schoenfeld, W. N. *Principles of psychology.* New York: Appleton-Century-Crofts, 1950.
15. Miller, N. E. Experimental studies of conflict. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* Vol. 1. New York: Ronald, 1944. Pp. 431–465.
16. Mowrer, O. H. *Learning theory and personality dynamics.* New York: Ronald, 1950.
17. Mowrer, O. H. Motivation. *Annu. Rev. Psychol.*, 1952, **3**, 419–438.
18. Mowrer, O. H., & Kluckhohn, C. Dynamic theory of personality. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* Vol. 1. New York: Ronald, 1944. Pp. 69–135.
19. Mowrer, O. H., & Lamoreaux, R. R. Avoidance conditioning and signal duration—a study of secondary motivation and reward. *Psychol. Monogr.*, 1942, **54**, No. 5 (Whole No. 247).
20. Mowrer, O. H., & Lamoreaux, R. R. Fear as an intervening variable in avoidance conditioning. *J. comp. Psychol.*, 1946, **39**, 29–50.
21. Mowrer, O. H., & Lamoreaux, R. R. Conditioning and conditionality (discrimination). *Psychol. Rev.*, 1951, **58**, 196–212.
22. Notterman, J. M. The interrelationships among aperiodic reinforcement, discrimination learning, and secondary reinforcement. *J. exp. Psychol.*, 1951, **41**, 161–169.
23. Page, H. A., & Hall, J. F. Experimental extinction as a function of the prevention of a response. *J. comp. physiol. Psychol.*, 1953, **46**, 33–34.
24. Schoenfeld, W. N. An experimental approach to anxiety, escape, and avoidance behavior. In P. J. Hoch & J. Zubin (Eds.), *Anxiety.* New York: Grune & Stratton, 1950. Pp. 70–99.
25. Schoenfeld, W. N., & Antonitis, J. J. A function of respondents in the extinction of operant responses. *Conf. exp. Anal. Behav.—Notes*, 1949, No. 17. (Mimeo.)
26. Sheffield, F. D. Avoidance training and the contiguity principle. *J. comp. physiol. Psychol.*, 1948, **41**, 165–177.
27. Sidman, M. Avoidance conditioning with brief shock and no exteroceptive warning signal. *Science*, 1953, **118**, 157–158.
28. Sidman, M. Two temporal parameters of the maintenance of avoidance behavior by the white rat. *J. comp. physiol. Psychol.*, 1953, **46**, 253–261.

29. SKINNER, B. F. The rate of establishment of a discrimination. *J. gen. Psychol.*, 1933, **9**, 302–350.
30. SKINNER, B. F. *The behavior of organisms.* New York: Appleton-Century, 1938.
31. SKINNER, B. F. *Science and human behavior.* New York: Macmillan, 1953.
32. STONE, G. R. The effect of negative incentives in serial learning: II. Incentive intensity and response variability. *J. gen. Psychol.*, 1950, **42**, 179–224.
33. THORNDIKE, E. L. *The fundamentals of learning.* New York: Teachers Coll., 1932.
34. THORNDIKE, E. L. *The psychology of wants, interests, and attitudes.* New York: Appleton-Century, 1935.
35. TOLCOTT, M. A. Conflict: a study of some interactions between appetite and aversion in the white rat. *Genet. Psychol. Monogr.*, 1948, **38**, 83–142.
36. WARDEN, C. J., & DIAMOND, S. A preliminary study of the effect of delayed punishment on learning in the white rat. *J. genet. Psychol.*, 1931, **39**, 455–461.
37. WARNER, L. H. The association span of the white rat. *J. genet. Psychol.*, 1932, **41**, 57–90.
38. WINNICK, WILMA A. A study of incipient movements in avoidance. Unpublished doctor's dissertation, Columbia Univer., 1950.
39. WYCKOFF, L. B. The role of observing responses in discrimination learning. Unpublished doctor's dissertation, Indiana Univer., 1951.

(Received February 4, 1953)

Psychological Review
Vol. 61, No. 1, 1954

THE MEASUREMENT OF VALUES[1]

L. L. THURSTONE

University of North Carolina

In this paper I shall try to summarize briefly the attempts of several investigators to extend the concepts of measurement to the subjective domain. While this work is admittedly crude and exploratory, the results do look promising so that this field should be challenging for further study. Here we shall give only brief statements of the fundamental ideas without details of theory or experimental procedure. Our purpose here is only to sketch the nature of this field of research.

When we propose to measure human values, colleagues in the humanities may shudder at the very idea. When I wrote a paper entitled "Attitudes Can Be Measured," some of my colleagues did shudder. They were sure that social attitudes contain some essence that could not be identified and measured. They were sure that, in making the attempt, we would measure only the trivial.

Human values are essentially subjective. They can certainly not be adequately represented by physical objects. Their intensities or magnitudes cannot be represented by physical measurement. At the very start we are faced with the problem of establishing a subjective metric. This is the central theme in modern psychophysics in its many applications to the measurement of social values, moral values, and esthetic values. Exactly the same problem reappears in the measurement of utility in economics.

In order to establish a subjective metric we must have a subjective unit of measurement. Before we can accept a subjective metric, it must satisfy the logical requirements of measurement as distinguished from rank order. These objectives have been approximated in the equation of comparative judgment and its variants.

Before proceeding to discuss the many applications of the subjective metric, we shall review briefly the principal psychophysical concepts by which a subjective metric can be established.

Let us consider these concepts in terms of a rather simple example, namely, the judgment of excellence of handwriting. When we look at several specimens of handwriting, it is fairly easy to select some that are considered to be excellent and others that are judged to be poor. In general, there is good agreement in such judgments. If we were asked to equate our judgments of excellence in a handwriting specimen to some physical measurements on the script, we would find it difficult. One of the main requirements of a truly subjective metric is that it shall be entirely independent of all physical measurement. In freeing ourselves completely from physical measurement, we are also free to experiment with esthetic objects and with many other types of stimuli to which there does not correspond any known physical measurement.

If we present a single handwriting specimen to a subject with the request that he tell us how good he thinks it is, then he must try to convey the degree of excellence in terms of words. It is well known that people vary tremendously in their use of superlatives in appraisals of experience, and, consequently, it is preferable to avoid such

[1] This paper was read at the Southern Society of Philosophy and Psychology in Knoxville, Tennessee, on April 11, 1952.

a direct procedure. Next we proceed to pairs of stimuli. We can ask the subject to judge which is the better of two specimens. In so doing, the subject gives his comparative judgment for each pair and he is not asked to give any verbal description of excellence.

The degree of excellence of a handwriting specimen is experienced by the subject in terms of some subjective process or quale. Since nothing is known about the neurological correlates of judgments of excellence of handwriting, we shall dodge all such terminology by merely referring to the discriminal processes by which the subject does, in fact, discriminate between the different specimens. These processes may be assumed to be physical or truly subjective according to the preferences of the investigator. His preference on this point has nothing to do with the subsequent development of the law of comparative judgment.

When the subject makes a judgment that one specimen seems to him to be better than another specimen, we postulate discriminal processes which differ in some manner in terms of which the percipient does make the discrimination. The more excellent specimen has some quale which differs from that of the poorer specimen. Imagine that the discriminal processes which correspond to different values are arranged in a spectrum from those discriminal processes in terms of which the percipient experiences the good specimens to the other end of the spectrum with discriminal processes in terms of which he experiences what he calls the poorer specimens.

Consider next the phenomena of dispersion. If one subject were to examine the same specimen in comparative situations on a large number of occasions, it is not to be expected that he would always experience a particular specimen with the same discriminal process. It can be assumed that the same specimen will be experienced in terms of discriminal processes in the same general region of the subjective continuum that has been postulated. So far we have no metric.

At this point we recall one of the fundamental restrictions on the problem of establishing a subjective metric. The discriminal processes must be assumed to be of such a character that they do not necessarily have intensities or magnitudes which can be in any sense measured. This is an old problem that was discussed many years ago in psychophysical theory. For theoretical considerations, imagine that the discriminal processes could actually be identified on each occasion when the subject makes a comparative judgment. The repeated observations of the same specimen can be assumed to produce an error variation from one occasion to the next. If we consider the relative frequencies of these discriminal processes as responses to the same stimulus, then we can postulate a Gaussian error distribution for the responses to the repeated observations of the same stimulus. Let us now assume that the spectrum of discriminal processes is stretched or contracted in different parts in such a way that the frequency distribution of these processes is Gaussian in terms of any given stimulus. Now we have a metric, but it is so far an entirely arbitrary metric. Imagine, at least in theory, that the same procedure can be repeated for many different stimuli which cover the whole range of discriminal processes in terms of which degree of excellence is experienced. It is now a question of experimental fact whether the metrics determined for the separate stimuli will be the same when all of the stimuli are considered together. It has been found in many experiments that such is the case.

If we represent in the same model the

comparative judgment of two stimuli in which the subject says for each presentation which of the pair is the better, then we can observe the proportion of attempts in which the subject judges specimen *j* to be better than specimen *k*. If we have a whole table of such proportions, it is possible to infer the spatial separations of the different distributions of discriminal processes. Each stimulus is then assumed to project a Gaussian distribution on the subjective continuum with a mean and a discriminal dispersion. An ambiguous stimulus will project a wider dispersion on the subjective continuum than a sharply defined or relatively unambiguous stimulus. Each stimulus will then be defined in the subjective continuum by its mean position which is called a scale value and by the standard deviation of its dispersion of discriminal processes. Each stimulus is then defined by two parameters in the subjective continuum.

Before we can put numbers into these parameters, we must define an arbitrary origin which may be taken as the mean value that one of the stimuli projects on the continuum. As a unit of measurement we may choose arbitrarily the standard deviation of the dispersion which that stimulus projects on the subjective continuum. When that has been done, similar numerical values can be assigned to all of the other specimens that have entered into the comparative judgments. Further, we can test for the internal consistency of this theoretical model.

It should be carefully noted that we have not assumed that the discriminal processes have magnitudes of any kind. They have been dealt with merely as subjective quales and we have assumed only that in principle their relative frequency of association with any given stimulus can be ascertained. While this cannot be done directly, these frequencies can be inferred indirectly from the observed comparative data. It should also be noted that we have not postulated the existence of any physical measures of any kind for the stimuli that have entered into the comparative judgments.

With this formulation of the law of comparative judgment, we are free to proceed with comparative studies of all kinds of stimuli which have no physical measure whatever. Hence we can turn to a wide array of interesting psychological problems involving value judgments. The freedom from any postulated physical measurement is the key that makes studies of this kind possible.

The method of comparative judgment turns out to be a rather general experimental procedure, and the well-known constant method in psychophysics is a special case in which one of the stimuli is arbitrarily taken as the standard which is compared with all of the other stimuli. Classical psychophysics was concerned with the more restricted problem of limen determinations.

We turn next to a brief review of some of the classical psychophysical methods because some of them have application in modern problems which transcend the determination of limens. In the method of equal-appearing intervals, the subject is asked to sort a large number of stimuli into a specified number of successive categories, say six or eight or ten. He is instructed to sort them in such a way that the intervals represented by the categories seem to him to be equal. This method is useful for rough survey purposes, but it can be shown that, even when the subject attempts to do this, he actually does not succeed in making the intervals subjectively equal. The method is, however, useful for coarse scaling such as the construction of attitude scales. The old method of equal-appearing intervals has been modified into what we call the method of successive intervals, in which the intervals are defined by descriptive phrases or by sample specimens. This method has been found to be very useful in various types of surveys to be discussed.

One of the old psychophysical methods was to ask the subject to sort a number of specimens into rank order. It has been found that rank orders can be analyzed in such a way as to obtain data approximately equivalent to that of the method of paired comparison. The method of successive intervals can even be analyzed as a variant of the method of single stimuli.

Since Weber's law and Fechner's law have figured so prominently in the history of psychophysics, we shall make a few comments about these two laws in relation to the modern setting. These two laws are frequently referred to as the Weber-Fechner law with the implication that they are the same law, but that is an error. It is possible to set up experiments with rather simple stimuli in which one of these laws will be verified when the other one is not verified. It would be useful to set up such experiments in order to show clearly the separation between the two laws. Weber's law states that the proportion of judgments $R > kR$ is a constant. R signifies here the physical magnitude of the stimulus and k represents another constant. Weber's law is concerned solely with physical measurements. It does not explicitly refer to the subjective continuum. On the other hand, Fechner's law states frankly the relation between the subjective continuum and the physical stimulus continuum. Fechner's law states that this relation is generally logarithmic, and it should be taken as a rough approximation to the relation between the subjective and the physical continua. Further, it can be seen that Fechner's law is applicable only to those stimuli which have a physical magnitude as well as an experienced intensity. The law of comparative judgment is completely independent of any physical stimulus magnitudes. The problem of the stimulus error is not ordinarily of serious concern to our problem. It deals with the ambiguity in the mind of the subject when he is asked to judge a stimulus as to the intensity of the subjective experience. Sometimes he attempts instead to judge the physical magnitude. A good example is that of a grocery clerk who can judge the weight of a bag of sugar. If he were asked to serve as a subject in the method of mean gradation, he would probably commit what Titchener would have called the stimulus error. In the measurement of social values, we are not interested in physical measurements because in general they do not exist for such values.

A very important advance in the application of psychophysical methods was accomplished by Richardson when he devised the triad method for studying the dimensionality of a domain. Instead of asking a subject to judge whether one stimulus is x-er than some other stimulus where x is any specified attribute, he set up the discrimination experiment in such a way that no attribute was specified. In the method of triads, the subject would be shown three patches of color, for example, and he would be asked to indicate which is the odd one with the implication that the remaining pair is more alike than any other of the three pairs. In this way the subject can make judgments of the degree of similarity or difference without having any specified attribute. Data collected in this manner can be transformed into the equation of comparative judgment and the dimensionality of the domain can then be ascertained by the Young-Householder theorem. Such a method can be used experimentally to determine the dimensionality of the various sensory modalities.

Perhaps the best known application of these experimental methods for the study of values is in the measurement of social attitudes. The most sensitive experimental procedure is to present the subject with pairs about which he is asked to make certain judgments. For example, he may be presented with pairs of nationalities, and he may be asked to judge for each pair which he would rather associate with. That type of experiment has been carried out in several ways. The judgments that are made by the subject depend, of course, partly on his own preferences which are closely related to his own nationality, and the judgments are also determined by the nationalities that are judged. If two groups of subjects are asked to make judgments of this kind, one can say on the basis of objective evidence which of the two groups is more tolerant of other nationalities. At one extreme we would have people who are completely tolerant toward all nationalities. They would then also, of course, be completely indifferent about their own. Such people would have no national loyalty or identification. At the other extreme we would have people who are said to be strongly prejudiced or biased. They would have extreme loyalties to some nationalities and extreme dislikes for others. I doubt whether we should consider either of these two extremes to be ideal.

Some years ago Fred Eggan wrote a master's thesis in psychology before he went into the field of anthropology. In that master's thesis he wanted to know the effect of different forms of question with reference to nationalities. He had five different questions representing different degrees of intimacy. All five groups of subjects were given the same lists of pairs of nationalities, but there were different questions. One group had the question, Which of each pair of nationalities would you rather associate with? Another group had the same nationality lists, but they were given the question, Which would you rather have as a fellow student? Another group had the question, Which nationality would you rather have your sister marry? The proportions were superficially quite different, but the rank orders of the nationalities were essentially the same. In this case, we would probably find that the form of the question has a tremendous effect on the discriminal dispersion but relatively little effect on the order of the nationalities. The effectiveness of comparative judgment for studies of this type should be exploited further.

In studying the measurement of social attitudes, the attempt is sometimes made to validate such experiments in terms of overt behavior, but that is an error. Samuel A. Stouffer of Harvard wrote a doctor's dissertation some years ago at the University of Chicago on this problem. He investigated social attitudes by means of statement scales in reference to the prohibition issue. He obtained data about his subjects as to their actual behavior on prohibition. He found that there was pretty fair agreement between what the subjects said on the attitude scales and how they actually behaved. I should like to point out that, while such a comparison is of considerable interest, it is not a validation of the attitude scale. A man may be entirely consistent in what he says and in what he does about a controversial issue, and yet both of these indices may be dead wrong in reflecting his attitude. In order to determine a man's attitudes in the sense of affective disposition about a controversial issue, it will be necessary for his friends to ask him privately when he is free to speak his mind and when he is not likely to be quoted. His personal attitudes may or may not agree with what he says and what he does. Here again, attitudes are essentially subjective experiences which may or may not conform with overt action.

Another distinction in the study of social attitudes which is sometimes lost sight of is that the cognitive and the affective appraisals may be entirely independent. For example, a group of subjects may agree in their strong dislike of communism. Someone might give them an examination in order to show that the subjects actually do not know what they are talking about. That might very well be true, but the psychological fact is nevertheless inescapable that the affective attitudes may be strongly for or against a stimulus even if there is a great deal of confusion about its cognitive description.

The statement scale is not so sensitive as the paired-comparison procedure. It consists in a set of statements to which the subject responds by acceptance or rejection of each statement. In constructing such a scale, one presents a large number of statements to a group of subjects whose principal qualification is that they can read English. These subjects are asked to indicate for pairs of statements which represents the stronger attitude for or against *x*, where *x* represents the psychological object to which the attitude scale refers. For rough survey purposes, the attitude scales are useful.

An interesting application of these methods of studying values is to appraise the effects of propaganda. We made a large number of experiments on the effects of motion-picture films on the social attitudes of high school children. Statement scales and paired-comparison schedules of various kinds were given before and after the showing of a motion picture. By this method we were able to ascertain whether a given picture had a significant effect and in what direction it did affect the children's social attitudes.

The method has also been applied in the study of international tensions by noting newspaper editorials. In one of those investigations a study was made with Chinese and Japanese newspaper editorials concerning each other, and it was shown, by treating key statements from the newspaper editorials, that the tensions increased at a very great rate before the two countries were at war. Quincy Wright has suggested in his political science studies that such applications of psychophysical methods might be useful in studying international tensions before they become very marked.

An application of these subjective measurement methods which has not yet been made will be in the definition of the morale of a group. In general, the morale of a group is described by newspaper reporters and by others who mix their own value judgments with the characteristics of the group to be described. For scientific work we should have a definition of morale which is entirely independent of the value judgments of the observer. Such a definition could be stated in terms of the dispersions of all of the debatable issues within the group. Other applications would be in the comparison of cultural and nationality differences as to the values that are considered to be essential. It is unfortunate that most students of social psychology and political science are too descriptively minded to adapt the quantitative methods that may be available.

Let us turn next to the experimental study of moral values. We have carried out several experiments in which a group of subjects was given a list of offenses that were presented in pairs. For each pair the subjects were asked to indicate which of the pair they considered to be the more serious. On the basis of data of this kind and with the aid of the equation of comparative judgment, we ascertain the scale values and dispersions for these offenses. In one case we gave a group of high school students such a list of offenses and we determined the scale values and dispersions for these stimuli for three occasions. The first presentation was a day or two before they saw a film that described the life of a gambler. A few days after seeing the film they were given the second similar schedule. About six months later they were given the third schedule. The film described the life of a gambler and we wanted to know whether this film had an appreciable effect on the attitudes of the high school youngsters toward gambling. We found that they considered gambling to be a much more serious offense after seeing this film than they did before seeing the film. In a number of experiments of this type, we also found that the motion pictures had much more lasting effects than is ordinarily supposed. In many cases we found that only half of the effect of the film wore off in six months. It should be said, however, that these experiments were carried out in small towns in Illinois where the children do not see so many movies as in the large cities. We carried out a similar experiment in the Hyde Park High School in Chicago where the children were given free tickets to a movie at the Tower Theater, a few blocks away. There we found that the effect was very slight. Our interpretation was that one movie more or less for children in a large city high school makes very little difference in their attitudes. These methods of studying moral values could be used very effectively in the comparison of different groups in a large city. The groups might represent different nationality backgrounds and different religious backgrounds. It would be interesting to ascertain what these differences would be. Such social psychological studies would help us to understand the problems of the extremely heterogeneous populations in the large cities. In a similar manner we have investigated experimentally the summation effect in propaganda where the effect of a single stimulus does not show a statistically significant effect.

Another interesting field of application is in experimental semantics. It would be useful, for example, to have an index of affective intensity for adjectives in a dictionary. Two adjectives may be equivalent as to cognitive meaning and yet differ widely in affective meaning. The words famous and notorious might be examples. So are the words pleasant, gay, and hilarious. Such affective indices would be useful in translating a foreign language.

We turn now to another type of psychophysical problem. In the psychophysical methods that we have considered so far the main problem was to allocate each idea or object to a subjective continuum which may be unidimensional or multidimensional depending on the nature of the problem. In most problems it is unidimensional. For example, if we ask subjects to judge the relative seriousness of offenses, we are dealing frankly with a unidimensional continuum, even though the discriminations may take place in a multidimensional continuum. We have here an obverse psychophysical problem. Having determined the subjective space which describes a group of subjects as to their attitudes in some field, we now inquire whether we can predict in any way what these people will do. When we turn the psychophysical problem in this manner, we find some exceedingly interesting psychophysical theorems of a new kind. I shall give a few examples.

Consider two political candidates for an election. Let one of them have a wide dispersion on the affective continuum. By this we mean that some people are very enthusiastic about this candidate, whereas others actually hate him. Let the other candidate have the same average popularity, but assume that he has a narrow dispersion so that very few people are enthusiastic about him and very few people strongly dislike him. If these two candidates come to an election, we should expect them to split the vote evenly. However, the more variable of these two candidates might introduce a third candidate of approximately equal popularity and who also has a narrow dispersion. Then we would have three candidates, one with wide dispersion on the affective continuum, and two candidates of narrow dispersion, and all three of them would be equally popular on the average. In such a situation, the more variable of the candidates would draw half the votes and the other two candidates would get twenty-five per cent each. These proportions would be altered somewhat depending on intercorrelations between the attitudes toward the candidates, but the principle can be illustrated in the general case for zero correlation. This principle is no doubt well known among politicians, but I doubt whether any of them have ever thought of this principle as a psychophysical theorem.

Let us turn to another simple example from the field of market research. Consider a mail-order house or a retail store which carries a limited number of neckties. They desire to please the majority of their clientele. The manufacturers offer many hundreds or thousands of necktie patterns. If you turn to market research people with this problem, they may ascertain the 20 or 30 or perhaps 50 of the most popular designs, and they may suggest that these be the designs that should be carried. But that is the wrong answer. Suppose that several hundred necktie patterns were submitted to a sample of the clientele. With such records one could rather easily determine not only which patterns should be carried, but also the number of patterns that should be carried in order to satisfy a specified proportion of the clientele. We would start with the most popular design and set that aside to be included. In the sample population we would then eliminate all who chose that popular pattern. Then we would inquire about the most popular pattern in the remainder of the sample population. That pattern would be set aside as the second design to be accepted. Eliminating those who chose that pattern, we would ascertain the most popular pattern in the remainder of the sample population. Proceeding in this way, we would come to the point where an additional pattern would increase the selection by only a very small percentage of the population and that would be the time to stop. In such a procedure we could determine the number of patterns as well as the designs which should be used in order to satisfy a specified proportion of the clientele. The ordinary solution of selecting the most popular designs would lead to a situation where some customers are confused by having many patterns which are equally acceptable while other customers find nothing to please them. The maximum satisfaction will be derived by proceeding in some such way as I have outlined. There is nothing profound about this procedure, and yet it would probably be novel in market research. There are situations where problems of this sort can be of national importance. If it should be necessary to restrict the manufacture of civilian goods, then it might be important to encourage the manufacture of a limited number of designs for all sorts of things and to select those designs in such a manner as to please the majority of the civilian population. In this manner the psychophysical methods may be important in contributing toward national morale.

Recently we made an experiment on the prediction of choice with regard to menus. In this problem we were concerned with the simplification of psychophysical methods to the point where they would be practicable for survey purposes. The psychophysical methods of the laboratory are often too laborious to be used in practical surveys. It was decided to adapt the method of successive intervals for this problem. We presented a list of 40 foods on a successive interval schedule in which each subject was asked to indicate by a single checkmark his relative degree of like or dislike for each food item. There were nine short descriptive phrases which represented degrees of like and dislike for foods. This schedule of 40 items required less than five minutes for each of several hundred adult men subjects. In addition to this short survey schedule, we also presented them with 16 menus in which they were asked to indicate what they would be likely to choose from each menu. For example, there were four lists of desserts, several lists of entrees, other lists of vegetables, and the like. For each menu the subjects were asked merely to check which they would select from a given list. Vanilla ice cream occurred in several of the dessert menus. The proportion of the subjects who select vanilla ice cream for dessert depends, of course, in part on their relative like or dislike for this dessert, but the selections would also depend on the competing items in the dessert list. By the application of the method of successive intervals and some theorems in psychophysics, we predicted the proportion of the subjects who would select each one of the items and there were 56 such predictions. These predictions were based entirely on the short, five-minute schedule for the whole list of 40 foods. We compared these predictions with the actual choices that the subjects made when they were confronted with the actual menus. The agreement was remarkable. The maximum discrepancy was between 3 and 4 per cent with one conspicuous exception for a dichotomy, namely, roast beef and fried chicken. The ratings for these two items were both in the upper two categories and the discrepancy was there 8 per cent, which was probably due to the effect of coarse grouping. The experiment demonstrated quite adequately that the prediction of choice can be effectively made with very simple survey schedules if these schedules are properly analyzed.

Some of these experiments deal with rather trivial values while others deal with socially more important values, but our principal concern here is in the development of those scientific methods which can be adapted over a wide range of values whether they be socially important or trivial.

We turn next to the application of psychophysical theory to some experimental problems in economics. For a long time there has been considerable interest in the measurement of utility, but the measurements have generally been indirect. Psychologists have been able to measure utility experimentally for over two decades, but economists have not until very recently expressed interest in these methods. In the last few years there seems to have been a marked change in the attitude of economists to these problems. In principle, utilities can be measured for an individual subject, but it is easier experimentally to apply these methods to the measurement of utility for a group of subjects. Psychophysical theory lends itself well to a number of variations in the measurement of utility. For example, the utility of a purchase can be described as the algebraic sum of the utilities of the object and of the price. In this case, the utility of the object would presumably be positive, whereas the utility of the price would be negative. The question then arises about the location of a rational zero point for the scale of utility. An experiment is now in progress to demonstrate an experimental procedure for locating the zero point in the scale of utility. It seems reasonable that the prices of various competing objects should be checked with their utilities to ascertain for any specified population to what extent some objects are overpriced or underpriced. Survey methods are available for doing these things. In determining the zero point for the scale of utility, we are asking several hundred subjects to express their preferences among various objects that might be given to them as birthday presents. Each of these single objects will then be given a value on the scale of utility. In addition to these judgments, we also asked the subjects to make a number of different judgments. We asked them whether they would prefer to receive gifts A and B or C. In this case they must judge whether the satisfaction from A and B is greater or less than the anticipated satisfaction from the single birthday present C. By judgments of this sort we expect to be able to locate the zero point of utility because the sum of the affective values of A and B combined should equal the utilities for these two objects taken separately. Within the range of the experiment with a small number of different objects to be selected, an additive theorem can be assumed to hold reasonably well. Diminishing returns would probably not be noticeable within the choice of four or five different objects.

In making these adaptations of psychological measurement theory to economics, one naturally wonders whether economics could be developed as an experimental science. Although I am not an economist, it has seemed to me entirely feasible that economics should be developed as an experimental science. In discussing this question with some of my friends in economics, I find that they are divided. Some of them insist emphatically that economics can never be an experimental science, while others are equally certain that this is possible. As an example we might consider the indifference function in economic theory. An indifference curve can be considered as a curve showing the combinations of two commodities X and Y which have the same utility value. If the amounts of the two commodities are considered to be the x and y axes in a three-dimensional model, then utility can be considered as the ordinates which are perpendicular to the x-y plane. An indifference curve would then be a horizontal section parallel to the x-y plane which represents constant utility. For different values of utility we would then have sections at different elevations which give a family of indifference curves. It has been shown that these indifference curves can be determined experimentally. There are many situations of controlled economies where the shapes of these functions can be studied experimentally. Such situations are in occupied countries or in prisons and in other situations with central control of prices. By altering the price of a commodity, the changes in the indifference curves can be noted experimentally.

As a final example of the adaptation of psychophysical theory in the measurement of values, we shall consider the field of esthetics. If esthetics were to be regarded as a purely normative science, then we should expect the esthetic value of an object to be determined by its physical properties. Such an interpretation seems well-nigh hopeless. It seems much more fruitful to recognize that the esthetic value of an object is determined entirely by what goes on in the mind of the percipient. In this manner of looking at the problem we deal again with values that are subjective experiences and which may vary from one person to another and certainly from one culture to another. An esthetic object symbolizes human emotional experience and its resolution in a conceptual and abstract manner. Except in extreme cases the esthetic experience is not itself emotional. It is essentially an abstraction. There is nothing absolute about the value of an esthetic object. The esthetic value is determined by the experience and the attitudes of the observer.

Some time ago I attended a series of seminars on esthetics at the home of one of my colleagues. Most of the participants in that seminar were from the humanities and the arts. The seminars were devoted to discussions about the theory of esthetics. In some of those discussions it occurred to me that the question at issue could be treated as a question of experimental fact, and I ventured to suggest how the psychophysical methods could be adapted to obtain an empirical answer to the question at issue. It was an illuminating experience to discover that some of my friends in the humanities were hostile to the very idea of subjecting questions of esthetic theory to empirical inquiry. On one of those occasions a friend showed me a quotation from Aristotle that settled the matter for him. It was heresy when I suggested that we knew more about this problem than Aristotle. Artists are sometimes suspicious of the experimental study of artistic preferences, and perhaps with some reason. Sometimes experimental studies are made in esthetics when the investigator is interested in secondary effects rather than in the esthetic experience. On the other hand, I have found some artists who are very much interested in such inquiry. A friend who is a portrait painter frequently encouraged experimental studies of this kind at the Art Institute in Chicago. Unfortunately I have not been able to induce many students of psychology to study experimental esthetics.

In closing I should like to comment briefly on the social studies as science. It is unfortunate that the social studies have rather low prestige among the sciences. I believe that this is what we should expect because a large number of researchers in the social studies have not adopted the impartial, objective, and intellectual attitudes of science. Quite generally in these fields the writers argue for social action of some kind, about the right and wrong ways of life, about what is good and what is evil in the opinions of the writers, about the good and the bad names and categories for describing their political friends and enemies. It is still true that social scientists rather frequently fail to study social phenomena as science to identify the forces at work without name calling and without injecting their own value judgments into what they are describing. As long as social scientists fail to distinguish between propaganda and science they will have low prestige among the sciences.

Summary

This paper has been concerned with the problems of a subjective metric. Social studies do not need to be quantitative in order to qualify as science. Some of the most important experiments in science deal first of all with the description of basic phenomena in a qualitative way. It usually happens that quantitative methods appear with more intensive study. Here we have considered some exploratory attempts to establish a subjective metric for the measurement of values. I have not succeeded in persuading social science students about the fascinating challenge to develop their field as science. To do so, we must free ourselves from the impulse for social action which has no place here. We should avoid problems in which we have an axe to grind. As citizens we have the privilege and the duty to participate in political elections. But when we work as scientists we should be aloof from the issues of the moment and to the chatter of the market place. Only in scientific detachment and objectivity can we eventually be helpful in developing the social studies as science.

(Received April 8, 1953)

Psychological Review
Vol. 61, No. 1, 1954

A NEURAL MODEL FOR SIGN–GESTALT THEORY[1]

JAMES OLDS

Harvard University

Whether we like it or not, a theory of learning points two ways. In one direction it points to better experiments. In the other direction it points to a model that would reproduce the aspect of behavior which the theory is used to explain; it is the unfinished blueprint for such a model.

It is not so readily understood that the first pointing depends on the second; the theory must point to a model in order to point to better experiments. Quite often, because this is not understood, a further implication is overlooked, namely, the more nearly finished the blueprint, the better the experiments will be. I will try to justify this proposition briefly in the next paragraph, but first I would like to emphasize its main consequence for the present discussion. This is that "mechanical" or "neural" models are superior to merely "conceptual" ones because they do provide a more nearly finished blueprint. They tell us not only the type of relations that must occur, but the type of material in which these relations must occur, and how the relations can be built into this kind of material.

The advantage of the completely specified model or mechanism would be to allow synthetic reproduction of the phenomenon under investigation. Synthetic reproduction gives the ideal solution to the main scientific problem: it apportions the variance of the phenomenon under investigation to the various causal constituents with no variance left over and not one too many causal constituents. Thus, it selects from the multitude of conditions that surround any phenomenon precisely the complex ingredients that are necessary to produce the phenomenon. In so doing, it gives the basis for a descriptive language that will not be crowded with irrelevant concepts, nor lacking in crucial ones, but rather will have just one concept for each important variable and none left over.

This would be the advantage of a completely specified model; the nearer we approach the completely specified model, the more we approach these advantages. Thus, it is to our advantage to get more specifications into the unfinished blueprint for the model. I believe the further implication is that an approach toward a mechanical model will always be beneficial.

The Adequacy of the Model to the Data

A model may fail, however, in either of two directions. On the one hand, it may be so incompletely specified as to fail to provide an adequate descriptive language and to carve out crucial variables. On the other hand, it may be more or less completely specified, but fail to reproduce the phenomenon under investigation.

My contention is that Hull's model (2, 3) is more completely specified than Tolman's (6); in this sense Hull has the edge. Tolman, on the other hand,

[1] This paper is based on portions of a dissertation submitted to Harvard University in partial fulfillment of the requirements for the Ph.D. degree in social psychology. The work was supported by a Research Training Fellowship of the Social Science Research Council and by funds of the Laboratory of Social Relations at Harvard. The writer wishes to express his appreciation to Professor R. L. Solomon for his many helpful criticisms and suggestions.

presents a model that seems to reproduce more adequately the phenomena of learning and performance that are the subject matter of both theories; in this sense Tolman has the edge. I want to consolidate their gains.

My purpose in the present paper, therefore, is to set forth a more complete blueprint for the model which Tolman has presented. I will do this by giving a neural interpretation of Tolman's theory based in large part on Hebb's (1) discussion of the properties of cell assemblies.

ADVANTAGES OF THE MODEL

As the proof of the pudding must be in the eating and not in any complicated rationalization, I will suggest at the end of this paper some of the advantages produced by the additions which I make to the Tolman theory. These come under three headings: (*a*) resolution of the problem of latent learning, (*b*) the stimulus control of ideas, and (*c*) the growth of approach motives. As it would do no good to expand on advantages before we have the theory, we proceed immediately to an introduction of the various important points of the model.

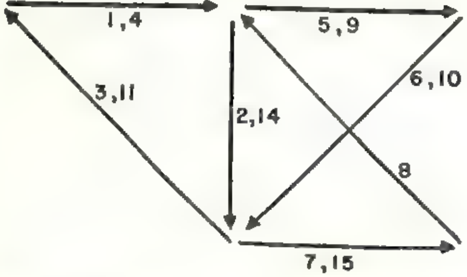

FIG. 1. The cell assembly described by Hebb (1, p. 73)

Hebb's Cell Assembly

The cell assembly described by Hebb (1) is most simply conceived as a three-dimensional lattice of neural paths providing several complete circuits, and alternative paths from each junction point so that when an impulse finds one of the transmission units refractory, another path allows the impulse to stay alive within the system. Therefore, the system has the capacity to reverberate. The assembly is most easily understood on the basis of the diagram in Fig. 1 borrowed from Hebb (1, p. 73). Each of the arrows in the diagram represents a single transmission unit, a single pathway. Although these are not considered by Hebb to be individual neurones, but rather low-order systems of neurones, we will take them to be the lowest order of functional units for our present explanation. Each pathway is refractory for a moment after an impulse has traversed it. Therefore, without alternative pathways reverberation would quickly die out, for the impulse would come back a second time before a pathway could recover. Each cell assembly consists in a number of these paths; the diagram represents a cell assembly. From the diagram, we can see how alternative pathways make reverberation possible. The impulse enters along the pathway marked 1,4, it proceeds to 2,14, and then through 3,11 and 1,4 again. At this point, it finds 2,14 refractory, but there is an alternative path, 5,9. The impulse proceeds around according to the numbers and is allowed to stay alive within the system because *neither all the pathways, nor too many of them are refractory at the same time.*

Hebb's cell assembly as it stands has five properties that we should note before we proceed. The first is reverberation. When an impulse enters the assembly it can reverberate without further stimulation. Second, the cell assembly has relations to other internal assemblies so that it can be aroused by central facilitation. Third, it has relations to the peripheral receptors so that it can be aroused by the environment. Fourth, it tends to have behavioral outlets, that is, it tends to control behaviors while it is aroused. Fifth, it has at least two states or phases: it can be latent when it is not aroused, and it can be in a state of reverberation when it is aroused.

The Four Phases of Ideas and Wants

At this point we turn to the aspects of behavior that are to be explained by Hebb's construct. There are two enigmatic terms avoided by S-R psychologists and often by cognitive psychologists because they seem so subjective and unfathomable, and so particularly refractory to mechanical analysis. These are "ideas" and "wants."[2] No psychology is lacking a set of euphemisms for these terms, but few psychologies handle the problems well. S-R psychology speaks of "fractional components" instead of "ideas," and of "antedating goal reactions" instead of "wants." Tolman faces the problem with less circumlocution: he speaks of the "expectancy" or the "significate" instead of speaking of the "idea." And he speaks of the "readiness" or the "demand" instead of the "want."

If, instead of searching for better and more satisfactory euphemisms, we take the terms as they stand with their more or less obvious, everyday meanings, and ask what we know about them, we find that we know quite a lot more than we might expect. And we also find that there is an interesting parallelism between an analysis of ideas and an analysis of wants that suggests that they are not such different things as they might seem at first glance.

The present analysis is going to be quite cursory and gross, for it is only to prepare the way for the model which is to come; it is to give some meaningful anchorage points for the technical material that is to follow.

First, there are various phases found in the analysis of a single idea. Let us take as an example the idea of a red light (of the traffic control variety). At first the red light is seven or eight blocks up the road, and we are not even thinking about it. I will say that the "idea of the red light" is *latent* at this point. After a few moments, we are approaching the intersection, the light at the corner turns from green to yellow, and for a very little while we are thinking about the red light and expecting it, but we are not seeing it. I will say the idea of the red light is now in a state of *expectancy*. But then the light turns red and we are seeing a red light. I will say the idea of the red light is in a state of *perception*. After we have sat behind the light for what seems like an interminable number of seconds, we become fed up with it, it seems to be lasting forever. I will say the idea is in a state of boredom. Finally, the light turns green, we drive on and forget it. The idea of the red light is latent again. It is obvious that an idea has at least four distinct conditions or phases: (*a*) it is not even thought, (*b*) it is thought but not seen, (*c*) it is seen, and (*d*) it is palling. The second condition can be divided again and again; when the idea is thought but not seen, it can be a mere thought, an expectancy, a memory, and so forth, but we will ignore these finer

[2] By the term "want" at this point, I refer to more than the basic physiological drives that underlie some (but not all) of behavior. Instead, I refer to the specific conceptualization of a goal that seems to precede most goal-directed activity in a human being.

gradations in the present paper. For our purposes, the idea has four phases which we may call latency, thought, perception, and boredom.

The most interesting thing about these four phases is that they are exactly and obviously paralleled by the four phases of a want. At first the want is latent; as for example when I am not thinking about food, and I do not want it. Next, something makes me think of food, and I notice that I am hungry. I start doing things that will get me fed; the want is now in a state of motivation. After that, I am being fed. The want is in a state of gratification. Finally, I am too full, and the want is in a state of satiation. After I have waited for a while, the satiation disappears, and the want is latent again. Thus, the want has four phases which we may call latency, motivation, gratification, and satiation. Note how closely these fit the phases of the idea.

From the parallelism, one would be tempted to suggest that ideas and wants are much the same sort of things. I suggest that they are not distinguished as far as the kind of structure is concerned, but only in terms of some power or "motive force" parameter. That is, an idea is a concept with a low motive force; a want is a concept with a high motive force.

The Four Phases of the Cell Assembly

The cell assembly, as we left it a few paragraphs back, has only two phases, latency and arousal. The state of arousal is a state of reverberation; an impulse enters the system along one pathway and reverberates within the system without further stimulus support.

We would like to find some characteristic of the cell assembly, implicit in Hebb's description of it, to allow us to ascribe it four phases and thus use it as an adequate model for the ideas and wants we have just described. Particularly, we would like to find two different conditions of arousal, one corresponding to perception or gratification and the other corresponding to thought or motivation. Analyzing these two phases, we find that the thought-motivation phase is characterized by a minimum of external stimulus support: it is a more or less autonomous internal reverberation, and it does not seem to be terminated either of its own accord or by mere withdrawal of the arousing stimulus. Rather, this phase of expectancy or motivation is terminated by the presentation of the goal object in the environment.

The perception-gratification phase, on the other hand, is characterized by a maximum of external stimulus support: it is not an autonomous internal reverberation, it does seem to become satiated or refractory of its own accord, and it seems to go out immediately upon withdrawal of the arousing stimulus. This phase is the perception or enjoyment that is turned on by the goal object in the environment.

Our problem is this: How can the same idea participate in an expectancy which is terminated by the goal object, and in a perception which is turned on by the goal object? The same idea seems to be turned off and on by the same object, which sounds ridiculous.

I find the answer to this question in Hebb's discussion of the conditions necessary for reverberation, an answer which shows that Hebb's cell assembly is a much better model for ideas and wants than one might expect from a superficial glance.

You will remember that in our description of the cell assembly, we said it would reverberate because *neither all the pathways, nor too many of them* *are refractory at the same time.* At the present point, this assertion becomes crucial. We may suggest that any stimulus which has a single or small number of connections with a given cell assembly would start a reverberation (a thought or motivation process in the assembly). A stimulus which has a large number of connections to many of the different pathways, on the other hand, would not set up a reverberation; instead it would "fire" the assembly. All pathways would be rendered refractory (or relatively refractory) at the same time. In the continued presence of the strong external stimulation the activity of the assembly could be maintained. But upon withdrawal of the external stimulus the assembly would be refractory, and activity would cease.

For our purposes, then, the cell assembly has four phases or conditions. We will say it can be in a state of latency, in a state of reverberation, in a state of firing, and in a state of refractoriness. These correspond to the four phases of ideas and wants. For the first phase we have used in all cases the term *latency*. For the second phase, we render equivalent the terms *thought*, *motivation*, and *reverberation*. For the third phase the equivalencies are *perception*, *gratification*, and *firing*. For the fourth phase the terms are *boredom*, *satiation*, and *refractoriness*. The cell assembly is our mechanical model for ideas and wants. A cell assembly of low "motive force" is an idea; a cell assembly of high "motive force" is a want. We will go on now to a simplified discussion of association.

The Association of Ideas

Again we turn to the aspect of behavior that is to be explained, and again we find a phenomenon which is rarely treated in contemporary psychology except with careful circumlocution. This is the association of ideas which is produced within a human being by a succession of stimuli in the environment. Each of us knows from his own experience a great deal about the way an associational link between two ideas functions, but we do not often analyze the functioning carefully enough to be aware of its essential characteristics.

I will take a simple example to make these characteristics explicit. Our subject is unacquainted with his typewriter. The carriage is far to the right, and he perceives and pushes a key marked "Tabular." The carriage jumps five spaces to the left and stops in a new position. First, there is an antecedent situation; then he makes a response and an outcome ensues. The antecedent situation is the carriage far to the right plus the perception of the tabular key; we will call this *A*. The response is to push the tabular key; we will call this R_1. The outcome is the carriage five spaces to the left; we will call this *B*. Thus, in the presence of *A*, R_1 leads to *B*. The *A*-R_1-*B* learning sequence has taught our subject that *A* followed by R_1 leads to *B*. We may say there is now an association of the *A* idea through R_1 to the *B* idea. In the future, if *B* is wanted, and *A* is presented, our subject will perform R_1. Also, if *A* is presented and R_1 should occur by accident, our subject will expect, and prepare for *B*. That is, if he wants the carriage moved from its *A* to its *B* position, he will now press the tabular key. And if he inadvertently presses the tabular key, he will expect and quite likely take some action to offset the movement of the carriage to its *B* position.

After a certain response in the presence of *A* has led to *B*, we say that some idea of *A* is associated with some idea of *B*. But the facts of behavior are these: (*a*) in the future if we make this particular response to *A* we will anticipate or expect *B*. (*b*) In the future if we should happen to want B we would show some tendency to search out A and then to make this particular response that takes us from A to B. The perception of A now arouses some expectancy of B, and the motivation of B induces motivation of A. The link seems to carry expectancy in the A to B direction, and motivation in the B to A direction. This will become clearer now as we lay out the specifications for our model in detail.

A NEURAL MODEL FOR SIGN-GESTALT THEORY

There are two undefined structural units of the model. These are the *cell assembly* and the *response control unit.* We presume at the outset that for any stimulus with which the subject has repeated commerce, a cell assembly becomes established; thereafter, the stimulus is an *unconditioned stimulus* of the cell assembly. Further, we presume that for any response which becomes organized within the behavior repertory of a subject, a response control unit becomes formed; thereafter, the response is elicited by the activation of the response control unit. These two formative processes may occur at first more or less by chance; the rules of organization and growth given below will show how selectivity can be introduced after a chance generation of these structural units.

In the exposition, cell assemblies will be designated by the lower-case letters of the early part of the alphabet, e.g., a, b, c. Response control units will be designated r_1, r_2, r_3, and so forth. Stimuli in the environment will be designated by the upper-case letters of the early part of the alphabet, e.g., A, B, C. Responses will be designated R_1, R_2, R_3, and so forth.

The definitions or specifications and postulates are listed below.

I. The unconditioned stimulus. Each cell assembly has a *stimulus threshold of firing* which needs to be crossed by stimulation from the environment. A stimulus which crosses this threshold is an *unconditioned stimulus* of the assembly. The unconditioned stimulus of assembly a is A, that of b is B, and so forth.

II. The conditioned stimulus. Each cell assembly has a *stimulus threshold of reverberation* which needs to be crossed by stimulation from the environment (mediated by antecedent assemblies as noted in VII below). A stimulus which crosses this threshold is called a *conditioned stimulus* of the assembly.

III. The motive threshold. Each cell assembly has a *motive threshold* which must be crossed by *combined positive motive force* or *combined negative motive force* (see VIIIa and b below).

IV. Intrinsic motive force. Each cell assembly has an *intrinsic positive motive force* and an *intrinsic negative motive force* which contribute toward combined positive and negative motive force respectively (and toward the combined motive forces of its antecedents when it is reverberating, see VIII below). Thus, there are two separate force parameters of each cell assembly; it is as though there were a solution with two separately variable factors dissolved.

V. The law of assembly activation. Both the motive threshold and one of the stimulus thresholds must be crossed at the same time for the assembly to become aroused (i.e., to fire or reverberate). If the motive threshold is crossed, then:

(a) the assembly will *fire* if the stimulus threshold of firing is crossed. Arousal ceases upon termination of this stimulus.

(b) the assembly will *reverberate* if the stimulus threshold of reverberation is crossed (unless both stimulus thresholds are crossed, in which case the assembly will fire). Reverberation continues after withdrawal of this stimulus; it is terminated by firing.

VI. The learning law of association. Two cell assemblies become related to one another by an associational relation under the following circumstances. If a fires, and then r_1 is activated, and then b fires, an associational relation will be formed between a and b which passes *through the response control unit* r_1. The cell assembly a will become the antecedent of the associational relation, and the cell assembly b will become the successor of the associational relation. They will be connected with one another through r_1. It is as though there were a wire connecting two terminal boxes a and b passing through a junction box r_1; and certain characteristics of the flow across the wire determine what will happen in the junction box (see IX below). The associational relation will be strengthened by further firings of a followed by activation of r_1 and firing of b. It will be weakened by further firings of a followed by activation of r_1 when these are not followed by firings of b.

VII. The law of conditioned stimuli. In the future, the firing of the antecedent will be a *conditioned stimulus* for the successor (see II and Vb above).

VIII. The law of the backflow of motive force. In the future, the reverberating of the successor will add two components of motive force, *instrumental positive motive force* and *instrumental negative motive force,* to the antecedent; these contribute toward respective combined positive and negative motive forces of the antecedent. A reverberating successor adds these components not only to the antecedent, but through the antecedent to further antecedents; the intervening assemblies need not be aroused for this transmission to continue to further antecedents.

(a) The *combined positive motive force* of an assembly is equal to the sum of its *intrinsic positive motive force* and its *instrumental positive motive force.* Similarly, *combined negative motive force* is equal to the sum of *intrinsic and instrumental negative motive force.* Either the combined positive motive force or the combined negative motive force of an assembly must be above the motive threshold in order for the assembly to become activated (see III above).

(b) The instrumental motive force (positive or negative) which a reverberating successor delivers to a near or distant antecedent is: (*i*) directly proportional to the combined motive force of the successor, (*ii*) directly proportional to the strength of the weakest link in the chain of associational relations between them, and (*iii*) inversely proportional to the number of assemblies interpolated between them.

IX. The law of performance. The likelihood of a response R_1 depends on the amount of *facilitation* and the amount of *inhibition* contributed to the response control unit r_1. Facilitation and inhibition are contributed to a response unit r_1 only when its antecedent a is firing and its successor b is reverberating. If a is firing and b is reverberating, then:

(a) Facilitation will be contributed to r_1 in proportion to the *amount of the difference* between the *combined motive force of the antecedent* and the *combined motive force of the successor* if this difference is *favorable to the successor.* Therefore, (*i*) if the successor is less negative than the antecedent, the response will be facilitated; (*ii*) if the successor is more positive than the antecedent, the response will be facilitated.

(b) Inhibition will be contributed to r_1 in proportion to the amount of the difference between the combined motive force of the antecedent and the combined motive force of the successor if this difference is *favorable to the antecedent*. Therefore, (*i*) if the successor is more negative than the antecedent, the response will be actively inhibited; (*ii*) if the successor is less positive than the antecedent, the response will be actively inhibited.

(*c*) If the facilitation is greater than the inhibition, then the response control unit r_1 will be activated, and R_1 will occur. If the inhibition is greater than the facilitation, then r_1 will not be activated, and R_1 will not occur.

X. The law of motive growth and decline. The intrinsic positive or negative motive force of an assembly grows and declines as a function of variables. I suggest the following postulates as a program for research.

(*a*) The intrinsic positive or negative motive force of an assembly is a joint, direct function of the number of transmission units in the assembly (see *b* below) and the amount of positive or negative motive force internal to each transmission unit (see *c*, *d*, *e* below). Each transmission unit has both positive and negative motive force internal to it.

(*b*) The number of transmission units in an assembly tends to *increase* in proportion to the amount of time the assembly spends in a state of firing.

(*c*) The amount of positive or negative motive force internal to each transmission unit in the assembly tends to *decrease* in proportion to the amount of time that the assembly spends in firing.

(*d*) The amount of positive or negative motive force internal to each transmission unit in the assembly tends to *increase* in proportion to the amount of time that the assembly spends in reverberation.

(*e*) The rate of positive or negative motive growth during reverberation (see *d* above) will *increase* as a function of the *combined* positive or negative motive force of the assembly during the period of reverberation.

(*f*) The rate of positive or negative motive decline during firing (see *c* above) will *decrease* as a function of the *intrinsic* positive or negative motive force of the assembly during the period of firing.

INTERPRETATION OF TOLMAN'S THEORY

We turn now to sign-gestalt theory to show that our mechanical model does give interpretation to all of its important points. We will first interpret the chief terms of Tolman's theory; then we will show how the relations postulated by Tolman are inferences from our model.

Perception: this is a term which is not accented as basic by Tolman; implicitly, however, it has a very basic place in his theory. For it is not the presence of a stimulus in the environment which controls behavior, in the Tolman formulation, but the "perception" of the stimulus by the subject. Perception is always selective; stimuli are perceived in proportion to their relevance to motives (6, p. 35). Tolman defines a perception as "an expectation of the component of a sign gestalt when this expectation results primarily from *present* stimuli coming then and there" (6, p. 452). I believe this may be paraphrased simply by saying a perception is the apprehension of an object by a subject when this apprehension depends on immediate stimulation. Our mechanical analogy for perception is the firing of a cell assembly. It requires both the presentation of the unconditioned stimulus (V*a*) and adequate combined motive force (V). The latter postulate accounts for the selectivity of perception.

Demand: this term is defined by Tolman as an "innate or acquired urge" to get to or from some given stimulus, or some physiological quiescence or disturbance (6, p. 441). Simply, this is a want; it is an appetite or an aversion. A demand in our mechanical system consists in either one of two states. In the appetite case, it consists in the reverberation of an assembly whose intrinsic positive motive force is sufficient to cross its own motive threshold; in this case, approach behavior will be elicited according to the law of the backflow of motive force (VIII) and according to the law of approach performance (IX*a*, *ii*). In the aversion case, it consists in the firing of an assembly whose negative motive force is sufficient to cross its own motive threshold. In this case, avoidance behavior will be determined jointly by the firing negative assembly and a less negative (or positive) reverberating successor. This determines behavior in the direction of the less negative successor according to the law of avoidance performance (IX*a*, *i*).

Sign-gestalt: this term is defined as the knowledge that a sign followed by a direction distance will lead to a significate, e.g., the knowledge that in the presence of *A*, R_1 leads to *B*. Our mechanical analogy for the sign-gestalt is two cell assemblies joined through a response control unit by an associational relation. The sign is the antecedent; the direction distance is the response control unit; the significate is the successor.

Sign-gestalt-expectation: this term refers to the expectation that a certain direction distance will lead to the significate; the expectation results from the fact that the sign is presented and perceived. Our mechanical analogy derives from the postulate that the firing of the antecedent arouses reverberation of the successor by the law of conditioned stimuli (VII). That is, if an associational relation joins *a* and *b* through r_1, then *a*'s firing arouses reverberation (expectation) of *b*.

Sign-gestalt-readiness: this term refers to a want for some means object by virtue of its instrumental relation to a demanded object. Our mechanical analogy here is the reverberation of a cell assembly whose intrinsic motive force is not sufficient to cross its own motive threshold. It requires reverberation of a successor (VIII) and presentation of the conditioned stimulus of the assembly in question (V*b*). The reverberating successor will add a component of motive force to the assembly in question; thus the combined motive force of the assembly will be above threshold, and the conditioned stimulus will arouse reverberation. At this point the assembly in question will function as though it were a "demand." However, termination of its reverberating successor will terminate its own demand characteristics, as its instrumental motive force supply will be cut off.

Sign-gestalt learning: Tolman's theory of learning is briefly the following. In any given training sequence, the subject learns new sign-gestalts, depending on what he perceives. For example, first the animal is in the presence of stimulus *A*. On Tolman's theorem of the selectivity of perception, the subject will perceive *A* provided that it is relevant to some present demand (6, pp. 35 and 386). Second, the subject adopts a direction distance R_2; that is, he performs behavior R_2. Third, when the behavior is done, he is in the presence of stimulus *B*. He will perceive *B* provided that it too is relevant to some one of his present motives. If the subject has perceived both the antecedent *A* and the outcome *B*, then a new sign-gestalt is learned in the performance process; it is that in the presence of *A*, R_1 leads to *B*.

Implicit in this description of sign-gestalt learning there is a premise that comes into superficial conflict with Tolman's (6, pp. 343–344) attack on the law of effect. The point is this: if the outcome B must be perceived in order for learning to occur, and if perception is contingent on motivational relevance, it follows that the outcome B must be either a goal or an instrumentality, a reinforcer or a secondary reinforcer, in order for learning to occur. But Tolman's attack on the law of effect suggests that possibly there is no need of B being a reward for learning to occur (6, p. 343). In justice we must say that Tolman (6, pp. 386–387) recognizes this superficial conflict, but he does not explicitly resolve the confusion. Our mechanical model does, and thus it provides a basis for reorienting the so-called "latent-learning" controversy (see Thistlethwaite, 5) as we will show in a moment.

Learning Requires Reinforcement

Our mechanical analogy for sign-gestalt learning derives from the learning law of association (VI). Two assemblies become related by an associational relation if a fires, then r_1 is activated, then b fires. But the conditions for the firing of a and b are outlined in the law of assembly activation (V). Both the motive threshold and the stimulus threshold of firing must be crossed before firing will occur. But in order for the motive threshold to be crossed, the assembly must have either sufficient intrinsic motive force (in which case its stimulus is a reinforcer) or sufficient instrumental motive force (in which case its stimulus is a secondary reinforcer). Thus, there is no learning without reinforcement.

But our model does predict *latent learning* provided the B stimulus *is* a reinforcing stimulus. For, a change in the combined motive force of b can be immediately reflected in two other changes: (a) a change in the combined motive force of a and (b) a change in the likelihood of the activation of r_1 while a is firing, both without any repetition of the A-R_1-B sequence. This derives from the law of the backflow of motive force (VIII) and from the law of performance (IX). The implication is that a change in the value of the outcome B will change the value of the antecedent A and the likelihood of the response R_1 to stimulus A without any repetition of the A-R_1-B sequence. Thus, learning which was *latent* when the combined motive force of b was insufficient to evoke performance will become evidenced when the combined motive force of b is changed by some operation.

Our suggestion vis-à-vis the rather large experimental program which has centered around the latent-learning controversy is this: experiments which succeed in making the outcome B sufficiently neutral with respect to the present motivational state of the subject will not give evidence of latent learning. We may just as well stop looking for learning without any positive or negative reinforcement, for in these cases the outcome will not be "perceived."

Experiments will demonstrate latent learning, however, whenever the outcome is made motivationally relevant in a positive or negative direction during learning, if the motivational relevance is reversed (as from positive to negative) after training without any further repetitions of the training sequence. In these cases, there will appear (if enough subjects are run) first-trial evidence of changes in response likelihood; such first-trial changes cannot be predicted by Hull's theory. Tolman and Gleitman (7) have reported such an experiment and it has sustained this prediction.

In summary, further experiments should show two things: (a) after A-R_1-B training with a reinforcing stimulus B, changes in the value of B will be reflected immediately in changes in the likelihood of the A-R_1 sequence without any further A-R_1-B sequences required to mediate this change in likelihood; but (*b*) learning will rarely be demonstrated in an A-R_1-B sequence where B has no history as a reinforcer or a secondary reinforcer, or where B is completely irrelevant to a strong present motivation, because in these cases B will not be perceived. In the terms of our model, b will not fire.

Stimulus Control of Ideas

The objection has long been made to cognitive theories that they do not genuinely predict behavior because they are unable to specify clearly before the fact the conditions under which the so-called immanent or ideational determinants of behavior will operate.

Our mechanical model for sign-gestalt theory takes a long step toward meeting this objection. The main cognitive determinants in Tolman's system are perceptions, expectations, readinesses, and demands. Tolman groups the first two, but we separate them. Our model specifies stimulus conditions, or operations under the control of the experimenter for the control of each of these cognitive processes.

Let us presume that our subject has been habituated to the sequence A-R_1-B-R_2-C-R_3-D. D is a primary goal, and thus this is the paradigm for any regularly repeated stimulus-response sequence eventuating in a goal. The internal organization resulting from the habituation will be a-r_1-b-r_2-c-r_3-d. To arouse the "perception of A" we must fulfill the conditions for the firing of a. Stimulus A plus some conditioned stimulus of d will suffice; for A is the unconditioned stimulus of a, and the reverberation of d assures the motivation of a. At the same time, we have fulfilled the conditions for the "expectation of B," that is, the reverberation of b. This is because a's firing provides a conditioned stimulus for b and d's reverberation provides adequate motivation; therefore b reverberates and B is expected. Although the conditions for the arousal of the "perception of A" are identical with those for the "expectation of B," the conditions for the termination of these two states are different. Firing of a will cease upon withdrawal of A; but reverberation of b will tend to continue until the presentation of B produces firing of b. Next, the presentation of a conditioned stimulus for d combined with a conditioned stimulus for c will produce a "readiness for C." This is because a conditioned stimulus combined with adequate motivation produces reverberation. The readiness will be terminated by presentation of C (which would fire c and thus terminate reverberation) or of D (which would cut off c's supply of instrumental motive force by terminating the reverberation of d). Finally, it is quite obvious that the presentation of a conditioned stimulus for d arouses a demand for D, and the presentation of D itself terminates that demand.

An experimental program which makes use of some of these specifications will be outlined briefly in the next section.

The Growth of Approach Motives

In conclusion, I am going to suggest briefly an experimental program for the investigation of the growth and decline of secondary approach motives based on the variables derived from the new model.

In the first place, it has been suggested that the intrinsic motive force of an assembly is a joint function of the number of "transmission units" in the assembly and the "motive force" vested in each unit (X*a*). The first problem in growing a motive, therefore, is to get some transmission units into the assembly, i.e., to get an assembly to start with. To do this we must give our subject some commerce with a stimulus, and then assure the firing of the newly formed assembly for some periods of time (X*b*). Presume that we want to form a motive directed at stimulus *B* as a goal. We may form an assembly and assure its firing by habituating our subject to the stimulus-response sequence A-R_1-B-R_2-C in which C is a primary goal. This forms the cell assembly *b*. We know the conditions for assuring the firing of *b*, namely, that during the time intervals while *B* is presented, if *c* is reverberating, *b* will be firing. During these periods of firing, *b* will be recruiting transmission units (X*b*) but these units will be losing motive force (X*c*). Thus, we are creating a cell assembly but not a motive.

In the future, however, the growth of positive motive force in *b* will be a joint function of time intervals of reverberation of *b* (X*d*) and the combined positive motive force of *b* during these intervals (X*e*), and the latter will be a function of the positive motive force in *c*, and the strength of the association between *b* and *c* (VIII*b*). To accomplish time intervals of reverberation in *b* we have to stretch out the time interval between A-R_1 and the presentation of *B*; that is, we have to give the conditioned stimulus which arouses reverberation in *b* and then delay the unconditioned stimulus which terminates this reverberation. Therefore, we delay the presentation of *B* with reference to its place in the habituation sequence. This delay should increase the intrinsic motive force in *b* (X*d*), and should result in a measurable increase in the reward value of the stimulus *B*. Increases in the reward value of *B* can be measured by changes in the subject's tendency to pursue this stimulus; I will not go into specific measures at this point, but they have been developed.

To accomplish a high combined motive force in *b* during intervals of reverberation, we have to assure a strong associational relation between *b* and *c*, and we have to make sure that *c* is reverberating during the delay. To vary combined motive force, then, we can vary the primary goal *C*, or vary the amount of habituation which establishes the associational relation.

In the future, the decline of positive motive force in *b* will be a similar joint function of time intervals of firing of *b* and the intrinsic motive force of *b* during those intervals of firing. The specific variables here are quite obvious, and I will not detail them here.

Experiments to carry out this program have been designed and some completed. Two experiments investigating motive force in *b* as a function of the delay of *B* have shown that after habituation this delay does produce significant motive growth (4). Experiments to test the effects of other variables are in progress.

SUMMARY

A mechanical model for sign-gestalt theory based on Hebb's (1) discussion of the cell assembly has been outlined. The cell assembly is used as the structural model for both "ideas" and "wants"; these two terms are rendered equivalent except that wants tend to have a higher motive force parameter than ideas. Cell assemblies have two kinds of activation, reverberation (corresponding to "thought" or "motivation") and firing (corresponding to "perception" or "gratification").

The model provides for the formation of associational relations among cell assemblies when there is a succession of stimuli in the environment. For example, if the objective stimulus-response sequence is A-R_1-B and so forth, where

A and B are stimuli, then an internal associational relation will be formed a-r_1-b, where a and b are cell assemblies, and r_1 a response control unit. After an associational relation has thus been formed between cell assemblies a and b through the response control unit r_1, the firing of a will tend to arouse reverberation in b, and reverberation in b (aroused from some other quarter) will add to the motive force of a and a's further antecedents. Thus, the associational relation passes stimulation forward from a to b and motivation backwards from b to a. Cell assemblies have two thresholds, a stimulus threshold and a motive threshold; both must be crossed simultaneously before any sort of activation will occur. The stimulus threshold may be crossed by either a "conditioned stimulus" (i.e., a firing antecedent) or an "unconditioned stimulus"; with adequate motivation, the former will produce reverberation, the latter will produce firing. The motive threshold must be crossed by the intrinsic motive force of the cell assembly or by a reverberating successor. Action is elicited when the antecedent assembly of a response control unit is firing, and the successor of the same relation is reverberating, and there is a motivational balance across the response favorable to the outcome.

The position adopted here represents an expansion of the position presented by Hebb (1). Hebb conceives facilitation as flowing both ways across an associational relation. However, he does not anywhere explicitly recognize the necessity that one particular kind of facilitation, namely, that which is here called motive force, can be conceived only as flowing from associational successor to associational antecedent if the problem of motivation is to be solved. I do not mean here that time in the central nervous system flows backwards. There is no hocus-pocus or magic here. My argument is simply that when cell assemblies are established in a communicating chain of circuits by the succession of their stimuli in the environment, then motivational flow will be from the representor of the successor to the representor of the antecedent.

The model is used to provide a reorientation of the latent-learning controversy. Latent learning is predicted in the sense that a change in the value of an outcome will change the likelihood of its preceding responses without further repetitions of the responses to mediate this change of likelihood. But the model fails to predict learning without reinforcement, for a stimulus must have value to be perceived (a cell assembly must have motivation in order to fire). On this basis, a change of focus in latent-learning experiments is suggested.

The model is used further to provide a new basis for research on the question of the functional autonomy of motives. Full-fledged learned drives are predicted, and the variables in their growth and decline are suggested. In general, it is suggested that the firing of an assembly increases the number of transmission units in the assembly, but decreases the motive force allocated to each transmission unit. Thus, it increases the size of the potential motivating unit, but decreases its motive force. Motive force, however, will grow later as a joint direct function of time intervals of reverberation, and instrumental value during those time intervals, and the size of the reverberating cell assembly. Firing will later tend to extinguish the motive force of an assembly.

The implication is that after habituation of a subject to a stimulus-response sequence such as A-R_1-B-R_2-C where A, B, C are stimuli, C being a primary reward, then the lengthening of the R_1-B time interval will tend to produce increments in the intrinsic reward value of the stimulus B, and lengthening of the time interval of presentation of B will tend to produce decrements in this intrinsic value. Experiments validating the first half of this generalization have been performed (4); others are in progress.

REFERENCES

1. HEBB, D. O. *The organization of behavior.* New York: Wiley, 1949.
2. HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
3. HULL, C. L. Behavior postulates and corollaries—1949. *Psychol. Rev.,* 1950, 57, 173–180.
4. OLDS, J. The influence of practice on the strength of secondary approach drives. *J. exp. Psychol.,* 1953, 46, 232–236.
5. THISTLETHWAITE, D. A critical review of latent learning and related experiments. *Psychol. Bull.,* 1951, 48, 97–129.
6. TOLMAN, E. C. *Purposive behavior in animals and men.* New York: Appleton-Century, 1932.
7. TOLMAN, E. C., & GLEITMAN, H. Studies in learning and motivation: I. Equal reinforcements in both end-boxes, followed by shock in one end-box. *J. exp. Psychol.,* 1949, 39, 810–819.

(Received April 8, 1953)

Psychological Review
Vol. 61, No. 1, 1954

THE PLACE OF PHYSIOLOGICAL CONSTRUCTS IN A GENETIC EXPLANATORY SYSTEM[1]

GUDMUND SMITH

University of Lund, Sweden

There are various ways of explaining behavior events physiologically. Let us distinguish here between (*a*) the use of physiological data, or of constructs derived from such data, and (*b*) the use of constructs which need not necessarily be verified under the microscope or in the EEG. Some of the more advanced psychological theories are based on hypothetical constructs, as, for example, Hebb's theory (4) and Klein and Krech's "conductivity" concept (5, 7). Such brain models seem to serve as substitutes for psychologically defined models partly because their units of analysis are easy to conceptualize, to handle. The present paper is, however, primarily concerned with the first, less sophisticated and more common kind of physiological theorizing in psychology indicating that physiological processes are *the* manifest reality underlying all behavior events. This approach has often been criticized, and we need not repeat the criticism here (8, 11, 14). Instead, the belief that physiological data represent the basis and origin of mental processes will be used here as a convenient starting point for further inquiry into the place and role of physiological constructs in psychology, especially in a genetic frame of reference.

The assumption that physiological facts represent a "basic level" in the individual, the last link in the explanation of mental processes, is part of a more general assumption that behavior data have to be referred directly to physical objects, inside or outside, in order to be understood at all. As suggested already by Natorp and Cassirer, however, psychology need not adopt this traditional method of physics and physiology but can (and should) adopt a method of its own. The aim of this method is not to make new constructs in the same objectivizing direction as the natural sciences, but to reconstruct physical objects by tracing them back to their origin, the experiencing subject. Instead of using hypostatized constructs, such as body structure and outside objects, as an explanatory basis for mental processes, the psychologist should analyze the constructs themselves with respect to their genesis in mental processes. Consequently, a physical-physiological unit might be regarded as the outcome of a more or less condensed series of behavior events (perception, concept formation, etc.), the early stages of which are the prerequisite for the later, more adapted and objectivized ones.

An explanatory model concerned with physical-physiological categories is a generalized, abstract conception of reality, in many respects the end product of the conceptual development of Western science. Similarly, a physicalistic ("reality-oriented") frame of reference accepted by the individual can be described as the result of a far-reaching emotional-intellectual socialization. Piaget and Rapaport, among many others, follow in detail this development from primary to secondary stages in our cognitive schemata and thought processes, this acceptance by degrees of a common, objective knowl-

[1] The author wishes to express his gratitude to Drs. George S. Klein and Daniel J. Levinson for valuable criticism.

edge, of detours in thinking (9, 10, 12). Let us, therefore, understand physiological constructs or facts as signs of a more or less objectivizing (physicalizing) set or point of view in human beings; let us regard their role as frames of reference for reality-testing in the individual's development, his adaptation to a stabilized world. The proposition is, then, that the physiological "reality" determines behavior, not merely as a number of causal factors behind the "mental surface" but as a conception in the individual himself of human nature, of reality.

Emmert's law, stating that the apparent size of an afterimage varies directly as to the subject's distance from the projection field, may serve as an illustration (13). According to the proposition, it can be assumed that these size relations hold true when the experienced world of an individual (his relevant region) is conceptualized in a physicalistic, "accurate" way. This implies that afterimage and screen must become isolated from each other, the afterimage as a "subjective" and the screen as an "objective" phenomenon. The afterimage, conceived of in this conventional, physicalistic way, is a constant nerve process; the screen is a nerve process changing in inverse proportion to the screen's distance from the eye. Naturally, the subject need not know anything about retinal areas and the like, only the formal differences between the stable (inside) and the changing (outside) reference systems. As pointed out in an earlier discussion on Emmert's law, the arrangements in most afterimage experiments of this type favor an isolation of image and screen, favor the analytic set necessary to diminish size constancy as far as possible (13).

This being true, the relations in the world we perceive must become equivalent with the relations in the physiological schema. When now the screen is moved to or from the subject the area of stimulated nervous tissue will be extended or diminished—in linear proportion to the distance—but the afterimage (as excited area) remains constant. Hence, the afterimage will be small or large, respectively, as compared with the excited area of the projection screen. And the same relations appear for the perceived world of our "objective" subjects; their afterimages conform to Emmert's law. But as soon as the conceptual schema is less developed, or, as soon as it is different, there will be deviations from the rule. Children, for instance, often think that the afterimage is a real object like the projection screen, i.e., an object the size of which varies in the same way as other external objects. Consequently, their afterimages do not increase or decrease in relation to the screen at various distances but are apparently size-constant (13). In many adults, too, a negative afterimage (or an eidetic image) is first considered to be an object "out there"; not until late in a series of experiments is the size constancy overcome.

Thus, the variations in apparent size of projected afterimages differ among people because the conceptual frames of reference adopted by them are different. While a physiologist would probably prefer to say that the afterimage follows Emmert's law because of an underlying, constant nerve-process (which might be unstable in children and some adults),[2] the more reasonable explanation, considering the deviations reported above, seems to be that the individual, for some reason or other, has adapted a conceptual schema in full agreement with a world of linear physical relationships. It is now easy to see

[2] The more advanced physiologizing psychologist would, of course, use a hypothetical variable to explain deviations in the law. This kind of theorizing will be discussed later.

why many a theory bound to manifest physical structures or observed physiological processes succeeds in explaining only specific and limited forms of behavior; and one understands why some physiological psychologists are eager to make "pure perception" the main object of a psychological science. The classical experimental psychology has sometimes been able to explain response only because the (conventional) physicalistic view is generally accepted in our society. Percepts can be considered as representations of behavior events within a more or less normalized framework of external reality, and, therefore, they must partly agree with a popular, physicalistic model of the world.

The assumption that people perceive (behave) according to conceptual patterns as developed in their life history is not new; indeed, it has been stressed by Jackson, Head, Gelb, Stern, Cassirer, and many of their contemporaries, and later by students of perception and personality (6). Studies of cultural factors in perception, as, for example, comparisons of Rorschach responses in Western communities and primitive tribes (2), also tend to support the assumption; religion, customs, prejudices, the whole reality imposed on us by our society seems to determine what we actually perceive.[3] Cases of brain-injured people perhaps illustrate the point most clearly. One of Gelb and Goldstein's subjects, for example, did not see a red color as red in general but only as a specific hue related to well-known objects (e.g., strawberry), because his approach was nonsymbolic, because his conceptual schemata lacked centers for a categorized perception of color. His vision in a narrow sense was not impaired, however; he was supposed still to have receptors for red "in general." But the subject himself could not accept this abstraction any more (3). Psychosomatic medicine can furnish us with further data, e.g., the acceptance of a somatic cause of mental troubles may result in somatic symptoms.

Before concluding this discussion let us develop the considerations once again, but now in terms familiar to the traditional psychologist. It is not necessary to avoid the stimulus-response model altogether in order to show why physical facts often fail to explain experience and behavior data. Stimuli (from outside) and physiological processes have been defined above as objective, generalized conceptions of reality as developed in the empirical tradition of natural science. In the stimulus-response model, behavior is influenced by external stimulation as well as internal (body physiology). But the response is not necessarily directly determined by this stimulation and coherent with its properties; it is, instead, an expression of *how* the stimulation has been received and "acknowledged." A behavior event will become the immediate reflection of a physical-physiological process only if this process is conceived of as "reality" by the subject. As soon as the individual's conception of reality is less "objective," less socialized, the response cannot and will not be a mere prolongation of stimulus (inside or outside). The response or behavior, defined as the outward expression of our experienced world or relevant region, has an immediate physical-physiological basis only when this world is a stimulus reality.

The physiological "level" does not represent the origin of a mental development but a stage in it, often (but not necessarily) the end result of the socialization process of thinking (cf. 4). It seems to be meaningless to ask for physiological facts underlying be-

[3] The developments in this field have been excellently summarized and commented on by Dennis (1).

havior phenomena of an individual without knowing whether or not he has accepted the generalized cognitive schema to which these facts belong. This might explain why neurological models as described in the introduction had to be extended over the traditional boundaries of a matter-of-fact science in order to cover more than limited forms of behavior. If, for instance, an individual persists in behaving abnormally in spite of the fact that all known neurological functions in him seem to be normal, if he refuses to adopt the neurologist's reality and is solely governed by his own "unsocialized" experience, it becomes necessary to introduce a hypothetical construct (e.g., integration of brain processes), the derivations of which should be able to explain all behavior deviations, even those without a known physical basis or with an imagined one. This means that neurological constructs in psychology must be more concerned with the reality represented by the wide developmental range of human experience (behavior) than with the limited reality of manifest physical facts and physiological observations, i.e., these hypothetical constructs must remain basically psychological in spite of the physiological language.

The empirical question is, however, how the generalized behavior has developed in different individuals, or why it has developed in some individuals but not in others. A physicalistic schema as accepted by the individual thus gets a personal significance; it may, for instance, be looked upon as a communication or defense mechanism, as a cognitive style, etc. (17). The physiological conception of the world, the impersonal behavior, like all behavior phenomena, ought to be genetically explained (16).

REFERENCES

1. DENNIS, W. Cultural and developmental factors in perception. In R. R. Blake & G. V. Ramsey (Eds.), *Perception: an approach to personality.* New York: Ronald, 1951. Pp. 148–169.
2. DU BOIS, CORA. *The people of Alor.* Minneapolis: Univer. of Minnesota Press, 1944.
3. GELB, A., & GOLDSTEIN, K. Psychologische Analysen hirnpatologischer Fälle. X: Ueber Farbennamenamnesie. *Psychol. Forsch.*, 1924, **6**, 127–186.
4. HEBB, D. *The organization of behavior.* New York: Wiley, 1949.
5. KESSEN, W., & KIMBLE, G. A. "Dynamic systems" and theory construction. *Psychol. Rev.*, 1952, **59**, 263–267.
6. KLEIN, G. S., & KRECH, D. The problem of personality and its theory. *J. Pers.*, 1951, **20**, 2–23.
7. KLEIN, G. S., & KRECH, D. Cortical conductivity in the brain-injured. *J. Pers.*, 1952, **21**, 118–148.
8. LEWIN, K. *A dynamic theory of personality.* New York: McGraw-Hill, 1935.
9. PIAGET, J. Principal factors determining intellectual evolution from childhood to adult life. In *Factors determining human behavior,* Harvard Tercentenary Publ. Cambridge: Harvard Univer. Press, 1937. Pp. 32–48.
10. PIAGET, J. *La naissance de l'intelligence chez l'enfant.* Neuchatel: Delachauz & Niestle, 1948.
11. PRATT, C. C. *The logic of modern psychology.* New York: Macmillan, 1939.
12. RAPAPORT, D. Toward a theory of thinking. In D. Rapaport (Ed.), *Organization and pathology of thought. Selected sources.* New York: Columbia Univer. Press, 1951. Pp. 689–730.
13. SMITH, G. *Psychological studies in twin differences. With reference to afterimage and eidetic phenomena as well as more general personality characteristics.* Lund, Sweden: Gleerup, 1949.
14. SMITH, G. *Interpretations of behavior sequences. With respect to a radical change in the objective situation.* Lund, Sweden: Gleerup, 1952.
15. SMITH, G. Sprache und Erlebnis. *Theoria,* 1952, **18**, No. 1, 78–86.
16. SMITH, G. Development as a psychological reference system. *Psychol. Rev.*, 1952, **59**, 363–369.
17. SMITH, G., & KLEIN, G. S. Cognitive controls in serial behavior patterns. *J. Pers.*, in press.

(Received April 8, 1953)

A NOTE ON STIMULUS INTENSITY DYNAMISM (*V*)

FRANK A. LOGAN

Institute of Human Relations, Yale University[1]

In the recent version of his theory (4), Hull postulates as an intervening variable, stimulus intensity dynamism (*V*), which is defined as a function of the intensity of the stimulus and which enters multiplicatively into the determination of excitatory potential. The choice of a theoretical assumption is, of course, the right of the theorist so long as useful predictions follow. However, this paper will attempt to show how Hull might have deduced the relevant empirical phenomena from his theory without the use of *V*.

There are four general data areas for which *V* was especially designed. Let us summarize these and then propose an alternative description.

The first area is the classical conditioning situation where, for example, an increase in illumination is followed by a UCS. If two groups of subjects are exposed to this situation, where all known relevant variables are identical with the exception of the intensity of the CS (i.e., the amount of increase in brightness) the probability of the CR is greater for the group with the more intense CS (e.g., 5).[2] Hull would deduce this result on the basis of the difference in *V* between the two groups.

Let us, however, recognize that, between trials, the subject is in the contextual environment (S_{ce}) containing a dimly illuminated disk, and that any occurrence of the response to S_{ce} is not reinforced by the UCS. When the organism is in the more brightly illuminated environment (S_1), the occurrence of the response is repeatedly rewarded. The situation becomes a discrimination problem in which reinforcement follows the response to S_1 but not to S_{ce}. For a second group of subjects, also nonreinforced for responding to S_{ce}, the rewarded stimulus complex is a still more brightly illuminated environment (S_2). It follows that, since the difference between S_{ce} and S_2 will be greater than the difference between S_{ce} and S_1 (assuming that similarity is a monotonic function of stimulus intensity), there will be greater generalization of the inhibition conditioned at S_{ce} to S_1 than to S_2. Hull has provided the derivation that the net discriminatory excitatory tendency (${}_{S}\dot{\underline{E}}_{R}$) will be greater at the positive stimulus the greater the difference between the two stimuli. Therefore ${}_{S}\dot{\underline{E}}_{R}$ will be greater for the group with S_2 as the CS than for the group with S_1; a greater probability of CR is expected at the stronger stimulus.

In this derivation, we assumed that the CS represented an increase in the intensity from a zero or minimal value. If, however, the CS were a decrease (so that, between trials, the illumination would be brighter than any stimulus value used), the analysis here presented would lead to the expectation that a group with the more intense CS (but a smaller change from the intertrial situation) would perform more poorly than a second group

[1] The writer is indebted to Drs. Mark A. May, Neal E. Miller, and Burton S. Rosner for a preliminary reading of the manuscript. A research project designed to test quantitatively several of the differential predictions herein contained is being supported by a grant from the National Science Foundation.

[2] This generalization is not unequivocally supported (e.g., 2). Kessen (5) has suggested a possible analysis of the conflicting results.

with a weaker CS. That is, the nonreinforced S_{ce} in the derivation would be at the upper end of the intensity continuum, and inhibition would generalize down toward the less intense values. It would also be possible to have the stimulus at some intermediate value between trials, and for one group to have a lower intensity serve as the CS, and for another group, a higher intensity. Stimulus intensities appropriately chosen as equal j.n.d. distances away from the intermediate stimulus should give the same probability of CR even though one is more intense than the other. The postulates containing *V* would be forced to deduce that the difference between the groups favoring the more intense CS would still obtain even under these diverse conditions.

The second set of data for which Hull has found it expedient to employ *V* concerns those experiments dealing with the time interval between the CS and the UCS. These data suggest that optimal conditioning will obtain at some asynchronism around one-half second, and that intervals either longer or shorter are less effective (e.g., 6). For this reason, Hull's system postulates a stimulus trace which changes as a function of time, and for which a molar stimulus equivalent is calculated for substitution in the equation for *V*. The trace represents a changing dynamism, and the level of conditioning is assumed to depend upon the *V* occasioned by a trace of the appropriate age.

The present position would suggest a somewhat different interpretation of the stimulus trace: the numerical value of the trace describes the degree to which a trace of that age represents a change from the conditions of stimulation prior to the onset of the stimulus. The trace function thus states that the onset of a stimulus produces a continuous change in the stimulus complex, rising rapidly to a maximal difference at about one-half second, and thereafter being reduced until the stimulus complex is, effectively, as it was prior to stimulation. The discrimination learning paradigm again argues that the degree of conditioning will be directly related to the difference between S_{ce} and the CS, where this difference is partially a function of the time since the onset of the CS.[3]

The third general class of empirical phenomena for which *V* is directly applicable refers to primary stimulus generalization along intensity as a continuum. Let S_{ce}, S_1, and S_2 be as above, and choose another stimulus intensity (S_3) which is (*a*) more intense than S_2, and (*b*) equal j.n.d. steps away from S_2 as is S_1. After the CR is established to S_2, generalized response tendency is obtained at S_1 and S_3. Assuming that equal j.n.d. separation means equal difference, habit generalization from S_2 should be the same to each of the two test stimuli. However, it is found that the response strength is greater at S_3 than at S_1 (e.g., 3) which Hull would deduce on the basis of the difference in *V*.

If the same conceptualization as presented above is followed, the original learning involves a discrimination between S_{ce} and S_2; the inhibition at the former will generalize not only to S_2 but to other stimulus intensities. Since we have assumed that similarity is a monotonic function of intensity, S_1 will be more similar to S_{ce} than will S_3; S_1 will therefore receive greater

[3] Hull's postulate of the stimulus trace (*s*) does not contain the intensity (*S*) of the stimulus, but only time (*t*) since its onset. The most useful interpretation is that *s* is a calculational device so that $V = f(S,t)$. The analysis offered in this paper would favor the postulate $s = f(S,t,S_{ce})$.

generalized inhibition from S_{co} than will S_3. Thus, although S_1 and S_3 will each receive equal generalized habit from S_2, ${}_S\dot{\underline{E}}_R$ will be greater at S_3 than at S_1 because there will be less generalized inhibition opposing it; a greater probability of CR is therefore expected at the stronger generalized stimulus.

According to the derivation given here, if the CS were a decrease in intensity (a strong intertrial stimulus), then greater generalized response strength should occur to a stimulus of weaker intensity than to a stimulus equally different from the original CS but stronger. Here, the implication of *V* is diametrically opposite.

The fourth and final general class of phenomena which involves the use of *V* occurs in a simple discrimination between S_1 and S_2 (used as above) obtained by the single presentation method. An organism is placed in a starting box and, shortly thereafter, a guillotine door is raised exposing a hinged door of either light or dark gray. This changes the stimulus complex into either S_1 or S_2, in only one of which locomotion through the door is rewarded. The response strength to the positive stimulus is found to be greater following the discrimination training if the more intense of the pair has been the positive stimulus (e.g., 1). Hull has derived this result on the basis of the greater *V* at the more intense stimulus.

If, however, S_{co} is also considered, it will be immediately seen that, when S_1 is the reinforced stimulus complex, both S_{co} and S_2 will be accruing inhibition which will generalize upon S_1 from both sides. When, however, S_2 is the positive stimulus, it will receive the same amount of inhibition generalized from S_1 as, in the reverse case, S_1 received from S_2; but the generalized inhibition from S_{co} will be less to S_2 as the positive stimulus than was the case to S_1 as the positive stimulus. Since S_2 would therefore receive the less total generalized inhibition were it the positive stimulus, ${}_S\dot{\underline{E}}_R$ would be greater when S_2 is positive than when S_1 is positive.

It should be possible to employ single presentation discrimination learning, but to insure that the subject never experiences the contextual environment of the stimulus except at times when either the positive or negative stimulus is present. This would preclude the development of inhibition of S_{co}. Under such conditions, the analysis followed here would deduce that ${}_S\dot{\underline{E}}_R$ would be identical whether the weaker or the more intense of the pair was the positive stimulus.

The derivations followed above have been more substantiative than exact on the assumption that anyone familiar with the theory will have sufficient facility with its application to discrimination learning to follow the sketch presented. It will be immediately apparent that this discrimination analysis leads to similar deductions as obtained by the use of *V* if three assumptions are fulfilled: (*a*) the subject is exposed to the contextual environment of the relevant stimulus, that (*b*) during such exposure there is a zero or minimal intensity of that relevant stimulus, and that (*c*) any performance of the response during these intertrial conditions is nonreinforced. Differential implications have been suggested if these assumptions are not met.

While the writer is not aware of research bearing directly upon these implications, several incidental findings seem to favor the present analysis. A number of experimenters have used the offset of a tone as a CS, obtaining satisfactory conditioning even though dynamism would be near zero. Since *V* is assumed to enter multiplicatively in determining excitatory potential, it would force $_SE_R$ to zero and predict no conditioning. Also it is common, though typically unreported, experience to observe the occurrence of the response between trials more frequently early in training than later. This would be consistent with the hypothesis that the response becomes extinguished to the contextual intertrial stimulus conditions.

Subsequent experimentation[4] may suggest, of course, that both the above analyses are necessary; that is, that there is an effect determined by the absolute intensity of the CS over and above the effect of the difference between the CS and the intertrial stimulus. More exact research is required before an adequate formulation can be stated.

[4] Subsequent to the preparation of this manuscript, Marvin Schwartz has obtained unpublished data suggesting that a weaker CS is more effective than a stronger one when the contextual intertrial stimulus is intense, and that the occurrence of the response between trials becomes less frequent with practice. The writer has also learned by personal communication that Dr. Charles C. Perkins, Jr. has independently obtained comparable results.

REFERENCES

1. ANTOINETTI, J. A. The effect of discrimination training upon generalization. Unpublished manuscript, 1950. (Quoted in Hull, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.)
2. GRANT, D. A., & SCHNEIDER, D. E. Intensity of the conditioned stimulus and strength of conditioning: II. The conditioned galvanic skin response to an auditory stimulus. *J. exp. Psychol.*, 1949, **39**, 35–40.
3. HOVLAND, C. I. The generalization of conditioned responses. II. The sensory generalization of conditioned responses with varying intensities of tone. *J. genet. Psychol.*, 1937, **51**, 279–291.
4. HULL, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.
5. KESSEN, W. Response strength as a function of conditioned stimulus intensity. Unpublished doctor's dissertation, Yale Univer., 1952.
6. REYNOLDS, B. The acquisition of a trace conditioned response as a function of the magnitude of the stimulus trace. *J. exp. Psychol.* 1945, **35**, 15–30.

(Received for early publication September 10, 1953)

VOL. 61, NO. 2 MARCH, 1954

THE PSYCHOLOGICAL REVIEW

THREE DIMENSIONS OF EMOTION[1]

HAROLD SCHLOSBERG

Brown University

All of you have had to face the problems in the general field of emotion, whether your interest was theoretical or practical. I think you will agree that the field is chaotic. When you try to organize it, perhaps for presentation in a course, you probably follow one of two obvious methods. You can admit that "emotion is only a chapter heading," to quote Madison Bentley; in this case you present a sort of smorgasbord of interesting and important facts, and then go on to clinical cases or to experiments on drives in white rats, depending on your inclination. Or, if you wish a more orderly presentation of the topic, you may build it around some of the many theoretical controversies that stud the history of the field. My preference is for the latter method, and after years of following it in class, I think I am beginning to get a satisfactory integration. Let me run over the major theories, and show how they come together.

The first controversy was between the James-Lange theory and that dictated by common sense. There seems to be little doubt that James hit upon an important truth, namely, that the responses one makes in an emotional situation are more than mere expressions of a mental state; to put a current term in James's mouth, *feedback* from skeletal and visceral responses is an important component of an emotion.

The next major controversy was between the James-Lange and Cannon-Bard theories. If we stop worrying about whether the alleged mental state, emotion, resides in the cortex or in the thalamus, much of this controversy is pointless. Indeed, we can combine the contributions of the two theories and say that the hypothalamus is the key integrating center for outgoing impulses, and also for the feedback impulses that James emphasized; in this sense, it may be the "center" for emotions as well as for drives. This brings up another theory, the motivational theory of emotions, but this theory didn't meet much resistance, for anyone with a feel for either the derivation of the word *emotion,* or for the analysis of behavior, must agree that emotion and drive are overlapping categories.

The most recent controversy was over the question of whether emotion is organizing or disorganizing. My answer to this question is, "Both." Love cer-

[1] This paper combines the material of two addresses: (*a*) "Some Dimensions of Emotion." Presidential Address, Div. 3, APA, Washington, Sept., 1952. Here the emphasis was on the qualitative dimensions, as found in facial expression. This material, in compressed form, appears near the end of the present paper, for the technical details of the study have since appeared in print (12). (*b*) "The Intensitive Dimension of Emotion." Vice-presidential Address, Sect. I (Psychology), AAAS, St. Louis, Dec., 1952. Read before a joint session of Sections I and Q (Education).

tainly disorganizes a student's study habits, but it does organize certain extracurricular pursuits! On the other hand, one of the most challenging tasks of the teacher is to arouse interest, a mild emotion, in apathetic students. Clearly one can't answer the general question; whether emotion is organizing or disorganizing depends on (*a*) the task under consideration, (*b*) the nature of the particular emotion, and (*c*) the strength of the emotion. Duffy (2) has long emphasized the last two points, at times suggesting that we stop talking about emotion, and substitute the direction of behavior, and the strength of behavior; for the latter she used the phrase "degree of energy mobilization." It is a troublesome term, for some people quibble about the meaning of *energy* in this context, and others worry about the implications of *mobilization*.

There are many other terms that may be used for this intensitive dimension. Cannon's concept of preparing for an emergency contains the same basic idea, but doesn't give us a good term. Perhaps the best name for the dimension is that used by Lindsley, "activation." His chapter in the recent handbook (9) gives us the outline of what he calls an activation theory of emotion. The term *activate* means a bit more than to make active; the dictionary tells us that it also means to make reactive. Activation would seem to be a very good name for what emotion does to us; the angry man overreacts to stimulation. Strong emotion thus represents one end of a continuum of activation; the other end, the condition of minimum activation, is found in the sleeping man who doesn't respond to stimulation. (If we wish to be accurate, we should put the state of zero activation at death, rather than sleep, for the sleeping man may respond to very strong stimuli. But psychologists don't study organisms at activation levels below that of deep sleep!)

To illustrate what we mean by the continuum of activation levels, let us start with a sleeping man, one near the zero level of activation. His cerebral cortex is relatively inactive, showing only slow bursts of electrical activity on the electroencephalograph. The muscles are relaxed and send few return impulses to the central nervous system. The sympathetic, or emergency, division of the autonomic nervous system is fairly inactive. As a result of this general condition, he doesn't respond to ordinary stimuli; he is unconscious.

Now let the alarm clock ring. It is a strong stimulus, and breaks through the high threshold. Gross muscular responses occur, and feed back impulses into the central nervous system. There is also autonomic discharge, and the resulting responses of muscle and gland lead to more feedback, probably through some interwoven pathways, the reticular substance. These impulses reach the hypothalamus, increasing its level of activity, and this center activates the cerebral cortex, as can be seen from brainwaves (9). In short, the individual is awake and responsive to stimulation. Perhaps I should have said, "more or less awake," for some individuals take a lot of time and activity, with resulting feedback, before the level of activation is high enough to permit anything but routine activities like dressing and lecturing!

Let us assume that our hero has reached an optimum level of activation by 10:00 A.M. He is alert, and responds efficiently to his environment. But now he finds that a book he needs is missing from his shelf. This frustration produces an increment in level of activation, perhaps not high enough at first to be dignified by the name of anger. But as he continues to search for the book the level of activation builds up until he is "blind with rage" or "functionally decorticate" to use Darrow's term; he probably wouldn't find the book now if it were under his nose. We will leave our hero in the range of level of activation that is conventionally set off as *emotion*, but let us not forget that he started at the other end of the continuum, sleep. The tendency to consider emotion as a separate state, divorced from the rest of the continuum, may well be the reason we have made so little progress in the field.[2]

Now let us consider level of activation in a more critical fashion: How are we going to measure it? Level of activation is a construct, crude at present. It is somewhat like level of prosperity. Everyone says that our country is more prosperous now than it was during the great depression. Suppose you wish a more precise statement, and consult an economist. He will quibble a bit, but ultimately he will probably suggest the use of a composite index, preferably based on key items like commercial bank deposits and payrolls of several large industries. Our economist will warn you that this gives an index of *general* level of prosperity, and that the level may differ in specific regions or industries.

Similarly, level of activation must be a general index, at least for the present. What are the key processes that we can use? There are a host of them, traditionally listed as bodily expressions of emotions. Blood pressure, heart rate, breathing indices, and hand steadiness are typical. A very promising one is tension in skeletal muscles, preferably accessory ones, as the brow potentials recorded by Kennedy and Travis (7, 14), or the neck muscles, with their important role in posture (10). But perhaps the most widely used measure is electrical skin resistance.

The psychogalvanic reflex, or better, the galvanic skin response, has been studied by hundreds of investigators. Perhaps its chief attraction is its sensitivity in mirroring ideational activity, particularly of an emotional nature. But its very sensitivity is largely responsible for the continuing argument as to whether or not GSR is a measure of emotion, for the response may be evoked quite readily by any sudden and strong stimulus, as a loud noise or an electric shock. Fortunately, the argument largely disappears if we drop the idea that emotion is a special state; strong stimuli, preparation to make an effort, and significant ideas all have a common feature, a quick increase in level of activation. This was recognized by Landis and Hunt (8) many years ago, when they stressed the fact that GSR is associated with an increase in subjective tension. As a matter of fact, tension is probably the popular word that comes closest to level of activation.

Consideration of the physiology of the GSR also points to its value as an index of level of activation. Darrow (1) showed that the fall in skin resistance was associated with the secretory activity of the sweat glands, under control of the sympathetic system. The skin resistance thus serves as an index of sympathetic discharge, a key element in the activation mechanism. He also pointed out that it is best to place the electrodes on the palms or soles, since these areas are relatively independent of thermoregulatory sweating. Finally, he suggested the best units in which to measure the phenomenon; rather than ohms, the usual measure of resistance, he preferred mhos, the re-

[2] Lindsley ends his chapter with the sentence, "In short, the activation theory appears to account for the extremes, but leaves intermediate and mixed states relatively unexplained as yet" (9, p. 509). The present discussion is less conservative.

ciprocal unit which describes conductance. This recommendation was based on the linear relationship he observed between conductance in mhos and the rate of sweat secretion. Later he suggested the use of log conductance, but a number of recent studies show that the mho is an excellent unit; it is normally distributed, and independent of the original level of resistance.[3] Conductance has the further practical advantage that it runs in the right direction, for increase in conductance is associated with increased level of activation, increased tension. Parenthetically, it is unfortunate that more workers don't calibrate their instruments in mhos, for it is at least questionable to perform even such a simple statistical operation as averaging on a skewed measure like the ohm.

The vast bulk of research on electrical skin conductance has lost the forest for the trees. Preoccupation with transient changes, the PGR, has led to general neglect of the slow drifts in absolute level of conductance, despite the fact that the absolute level is the obvious correlate of general level of tension or activation. This neglect is partly due to the design of conventional apparatus, which is adjusted to balance out basic level of conductance so that the transient changes may be read directly from deflections of a needle. For direct measures of levèl of conductance a much less sensitive and less elaborate apparatus is more adequate and convenient. I use a 50-microampere panel-type meter in series with a pair of dime-sized silver–silver chloride–saline paste electrodes. The potential to run the circuit is usually one volt, obtained from a flashlight cell with the aid of a potential divider and calibrated resistor; with this voltage, the dial needle reads directly in micromhos, and the range is adequate for most subjects. The whole gadget can be assembled for about $25, and is as portable as a box of cigars. It takes five minutes to attach the electrodes and adjust the apparatus, but only five seconds to obtain and jot down successive conductance readings. Thus, a single experimenter can carry on an experiment such as reaction time, taking periodic readings of conductance throughout the session. Although I haven't tried it yet, there seems to be no reason why this apparatus shouldn't be used to follow mean conductance level of groups of ten or more individuals engaged in a common task. For example, a pair of simple and reliable electrodes could be attached to each member of a small audience and connected in parallel to one meter. Since conductances in parallel summate, one would merely have to divide the total conductance by N to get the mean. Of course, small fluctuations, as asynchronous PGR's, would balance out, but the method should be perfectly adequate to determine the general changes in level of activation during the various episodes of a play, for example.

Now let me describe a few applications, to show that conductance serves as a satisfactory measure of level of activation.[4] One of the most convincing experiments is that of Duffy and Lacey (3). They recorded skin conductance on subjects who were going through several cycles of a psychophysical task. Conductance showed a sharp increase at the beginning of the first series of tones to be judged, and slowly fell during the progress of the series and ensuing rest period. Conductance shot up again at the start of the next series, dropped off during the series,

[3] See (13) for references.

[4] Skin conductance is not an ideal index of general level of activation. But short of a compound index, conductance may well be the best available, assuming that reasonable care is taken to keep electrode contact and room temperature fairly constant.

and so on. This saw-tooth pattern continued throughout the session, showing that the subjects alerted themselves each time they started a task, and then gradually relaxed as they made progress. Further, the general level of the conductance pattern fell from series to series within each session, and from day to day; the subjects were gradually relaxing as they became more familiar with the general situation.

Schlosberg and Stanley (13) have obtained results in an extensive series of tasks and tests run in five cycles over a two-hour session on each of five successive days. In addition to confirming the Duffy and Lacey findings, these experimenters hoped to relate conductance to efficiency. They ran into difficulties in the latter respect, for there was some suggestion of a curvilinear relationship between conductance and efficiency. This is what one would expect from what we said earlier about level of activation, for it seems likely that there is an optimal level of activation for each type of task, and perhaps for each subject. For example, a moderate level of activation would seem optimal for playing chess, whereas a relatively high one would be best for sprinting. In either case, the subject would report that he was too sleepy to do well if he were below the optimal level, and too tense if he were above it.

Plausible as this idea sounds, it is hard to pin down (10), for a lot of measurements are needed on each subject before we can obtain a good curve relating his efficiency to his level of activation. Freeman (4) showed one way to do it. He took short series of reaction times and simultaneously recorded skin resistances at various times during the day, depending on diurnal variations to give a broad range of levels of activation. His results showed a very clear inverted U relationship, with minimum reaction times at a moderate level of conductance. His published data are a bit scanty to establish such an important generalization, so it seemed desirable to repeat the experiment. I set up a portable reaction timer, with a built-in device to vary the foreperiod so that it could be used conveniently at home. A student has taken a hundred sets of readings on herself, sampling all hours of the day from before breakfast to bedtime. Each session included (*a*) conductance, (*b*) 20 simple auditory reaction times, and (*c*) hand steadiness. She obtained beautiful inverted U relationships between both hand steadiness and simple auditory reaction time on the ordinate and skin conductance on the abscissa. Her optimal level of conductance for hand steadiness is a trifle higher than that for reaction time. She also ran some short series of sessions on five other subjects, and they seemed to give comparable curves. These results are encouraging, for they seem to open the way to much fruitful work in the fields of skill, efficiency, and fatigue.

But you may feel that I have gotten quite far away from my title, "Three Dimensions of Emotion." Of course I have been dealing with the level of activation continuum, but perhaps you would like some studies on the high level of activation that is traditionally called *emotion*. I don't have anything very specific to report here, for emotions are hard to produce in the laboratory. I do have another student working on the effects of electric shock on conductance during a reaction-time task, but he has been too tenderhearted in adjusting the strength of the shock. However, I can at least point to the familiar lie detector test as a practical application of level of activation. The peak-of-tension method depends on a gradual increase in level of activation as the critical question approaches, followed by a marked fall in tension after the crisis has passed. Most interrogators prefer breathing and blood pressure as indices of tension, on the grounds that PGR is too sensitive (6). On the other hand, the PGR has been used successfully, and it seems probable that even better results would be obtained with a device designed to measure absolute level of conductance, rather than the quick swings.

The activation theory of emotions has one obvious failing: It deals only with the intensive dimension, and takes no account of differentiation among the various emotions. This is true as long as we limit our consideration to *general* level of activation, forgetting the fact that different subsystems might vary more or less independently in different emotions. The situation is quite analogous to that in the closely related field of motivation. Both hunger and thirst will raise the general level of drive, as measured on an activity wheel, but each will also act selectively on an appropriate family of S-R units. Unfortunately, the analogy isn't complete, for we haven't yet found differentiated emotional patterns as clean-cut as are eating and drinking. There are a few hints of such differentiation among bodily changes in the various emotions, but we need much more research before they can be established.

However, there is one field within the topic of emotions where we have long been embarrassed by the excessive number of different patterns: This is the field of facial expressions. Frois-Wittmann (5) brought some order out of this chaotic field by working out the interrelations among various expressions, and Woodworth (15) contributed a six-step scale that helped a lot. In 1941 Schlosberg (11) used this scale for collecting data on a new series of pictures, and found evidence that the scale described a roughly circular surface. Inspection of pictures arranged around the scale suggested that the surface might be generated by two axes, pleasantness-unpleasantness, and attention-rejection. The next step was to try to get a better description of the surface in terms of these two axes. In 1952 Schlosberg (12) reported the results of several attempts to obtain independent ratings on a large number of posed expressions, using nine-point rating scales. one for each dimension. Pleasantness-unpleasantness offered no trouble, but there was considerable difficulty in explaining the attention-rejection dimension to the judges. We tried pointing out that rejection was the active opposite of attention, characterized by compressed lips, nostrils, and eyes, as though forcibly excluding the external object, but this effort met with only mediocre success. We finally hit upon the use of "anchors"—pictures selected from another series to illustrate the extremes of attention and rejection. This stabilized the ratings and enabled us to locate each expression on the roughly circular surface described by the two dimensions, P-U and A-R. These positions were validated by using them to predict Woodworth scale judgments of the same pictures. The predicted Woodworth scale positions correlated with the obtained ones with coefficients of .92, .94, and .96 in three independent experiments, utilizing two different sets of photos of posed facial expressions. Hence, we may feel considerable confidence that P-U and A-R are two basic dimensions of facial expressions in particular, and perhaps of emotions in general.

We can compare the two dimensions of the facial-expression surface to the blue-yellow and red-green axes of the color surface. This immediately suggests that there may be a third dimension, corresponding to visual brightness. The third dimension for facial expressions might well be the intensitive one we considered earlier, level of activation. As a preliminary test of this possibility, we obtained ratings on the same pictures, this time using a rating scale that ran from sleep to tension. The results enable us to construct a crude three-dimensional figure, roughly comparable to the familiar Munsell color solid, and quite as irregular in shape (Fig. 1). The unpleasant pictures tend to show the highest levels of activation, with mirth at an intermediate level, while contempt, which combines pleasantness with rejection, has a rather low level of activation. The third dimension seems to clear up some expressions that are not separated by the original two axes; for example, grief, pain, and suffering all have the same P-U and A-R values, but grief is rated considerably below the other two expressions in level of activation.[5]

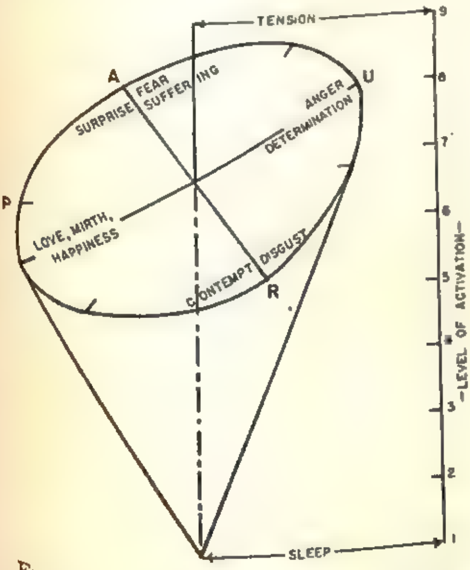

Fig. 1. A first approximation to the solid figure which represents the range of facial expressions. The emotions are placed correctly with respect to their maximum level of activation (sleep-tension) indicated on the ordinate. The top surface is sloped to show that anger and fear can reach higher levels of activation than can contempt. For a more accurate representation of the other two dimensions, pleasantness-unpleasantness and attention-rejection, see (12).

Much more work has to be done before we can be satisfied with the intensitive dimension of facial expressions. For one thing, we were working with a collection of pictures posed to represent emotions; this concentration on one end of the continuum introduces "series effects" in the ratings. We need a wider range of pictures, including low levels of activation such as sleep or listening to a dull lecture. Further, we should have actual skin conductance readings on the individuals, taken just before each picture was snapped. But these are projects for the future.

Summary

The activation theory of emotion brings together many of the theories and facts of emotion, at least as far as the intensitive dimension is concerned. Instead of treating emotion as a special state, differing qualitatively from other states, the theory locates emotional behavior on a continuum that includes *all* behavior. This continuum, general level of activation, has its low end in sleep, its middle ranges in alert attention, and its high end in the strong emotions.

Any one of a number of physiological processes may be taken as an index of general level of activation, but electrical skin conductance has certain advantages for the purpose. It is sensitive, easy to measure, and varies in a manner consistent with expected changes in level of activation. It promises to be equally useful in work on skills and efficiency, as well as on emotions.

Neither skin conductance nor any other physiological measure of level of

[5] The three dimensions attempt to describe pictures of the *responses* called facial expressions. Knowledge of the situation which evoked a given expression will help the judge to interpret the expression, but such situational cues need have no part in the description of the response per se.

activation has yet given us much beyond the intensitive dimension. Further research may furnish such evidence, but for the present we may profitably turn to facial expression to find the qualitative dimensions along which emotion may vary. Here, we have good evidence that the whole range of expressions may be described rather well in terms of a roughly circular surface, whose axes are pleasantness-unpleasantness and attention-rejection. We have some idea how level of activation comes into this figure as a third dimension, but further research is needed here, too.

Thus, facial expressions and body changes supplement each other in giving us the dimensions along which emotions may vary.

REFERENCES

1. Darrow, C. W. The significance of skin resistance in the light of its relation to the amount of perspiration. *J. gen. Psychol.*, 1934, **11**, 451–452.
2. Duffy, Elizabeth. The concept of energy mobilization. *Psychol. Rev.*, 1951, **58**, 30–40.
3. Duffy, Elizabeth, & Lacey, O. L. Adaptation in energy mobilization: changes in general level of palmar skin conductance. *J. exp. Psychol.*, 1946, **36**, 437–452.
4. Freeman, G. L. The relationship between performance level and bodily activity level. *J. exp. Psychol.*, 1940, **26**, 602–608.
5. Frois-Wittmann, J. F. The judgment of facial expression. *J. exp. Psychol.*, 1930, **13**, 113–151.
6. Inbau, F. E. *Lie detection and criminal interrogation.* Baltimore: Williams & Wilkins, 1942.
7. Kennedy, J. L., & Travis, R. C. Prediction of speed of performance by muscle action potentials. *Science*, 1947, **105**, 410–411.
8. Landis, C., & Hunt, W. A. The conscious correlates of the galvanic skin response. *J. exp. Psychol.*, 1935, **18**, 505–529.
9. Lindsley, D. B. Emotion. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 473–516.
10. Ryan, T. A., Cottrell, C. L., & Bitterman, M. E. Muscular tension as an index of effort: the effect of glare and other disturbances in visual work. *Amer. J. Psychol.*, 1950, **63**, 317–341.
11. Schlosberg, H. A scale for the judgment of facial expressions. *J. exp. Psychol.*, 1941, **29**, 497–510.
12. Schlosberg, H. The description of facial expressions in terms of two dimensions. *J. exp. Psychol.*, 1952, **44**, 229–237.
13. Schlosberg, H., & Stanley, W. C. A simple test of the normality of twenty-four distributions of electrical skin conductance. *Science*, 1953, **117**, 35–37.
14. Travis, R. C., & Kennedy, J. L. Prediction and control of alertness. III. Calibration of the alertness indicator and further results. *J. comp. physiol. Psychol.*, 1949, **42**, 45–57.
15. Woodworth, R. S. *Experimental psychology.* New York: Holt, 1938.

(Received March 22, 1953)

Psychological Review
Vol. 61, No. 2, 1954

A MATHEMATICAL MODEL AND AN ELECTRONIC MODEL FOR LEARNING[1]

L. BENJAMIN WYCKOFF, Jr.

University of Wisconsin

In a previous report (11), the author outlined a quantitative theory of learning which would take into account the learning of observing responses. This theory was left incomplete largely because a quantitative statement of secondary reinforcement was required, and no suitable statement was available. The present development has grown out of an attempt to fill this gap by postulating a quantitative statement of secondary reinforcement. The resulting scheme has turned out to be surprisingly simple. The inclusion of this postulate, rather than complicating the system, has actually simplified it, since it enabled us to discard certain other postulates with no apparent loss in explanatory power.

In spite of the simplicity of the basic postulates of the system, some difficulty was encountered in deriving the implications for experiments involving a number of stimuli and responses. In particular, it was difficult to determine whether the postulate system would actually yield plausible predictions of observing response learning. To facilitate this process a specialized analogue computer was constructed to operate according to the postulates. The device constitutes a robot which can be confronted with learning problems. Its performance corresponds to the performance prescribed by the postulates of the mathematical model, thus establishing the implications of the theory for the particular learning problem employed.

The device described here makes use of relatively elementary electronic principles and is not intended as a contribution to the engineering aspects of computer design. It is of interest because of the learning theory embodied in it and as an illustration of the way in which the actual construction of a physical model may facilitate theory development in psychology.

The electronic model was confronted with a discrimination problem in which it was required to "learn" to select, by means of observing responses, those aspects of the situation which were relevant, in addition to learning the correct choice in a two-choice situation. By this method we were able to test whether the present theory would imply observing response learning. In the first attempts. serious shortcomings of the theory were forcefully demonstrated. The machine did not learn the correct observing response, and in fact learned the opposite of what was expected. It was possible to revise the theory and the model with these failures in mind. The model described here is the corrected model. The original failure will be considered in the discussion.

Both the mathematical and electronic models make use of time intervals as a critical variable. Originally the mathematical model was based on probabilities, but because of technical considerations a modified model using time intervals was substituted. A close cor-

[1] Many of the ideas expressed in this paper were stimulated by discussions with N. Guttman and C. J. Burke to whom the author is indebted. N. Guttman and W. S. Verplanck have developed a theory of reinforcement which parallels the present model with respect to qualitative aspects. C. J. Burke has developed a statistical model which treats observing response learning.

respondence between the time interval and probability models exists and will be considered.

THE MATHEMATICAL MODEL

The central postulate of the present theory derives from a suggestion made by Skinner (10, p. 246) that discriminative stimuli tend to exhibit secondary reinforcing properties. In other words, if some response has been strongly conditioned to a particular stimulus, then this stimulus will tend to strengthen a new reflex upon which it is made contingent. This notion was further elaborated by Notterman (8), Schoenfeld *et al.* (9), and Dinsmoor (2). Experimental evidence was obtained which suggested that establishment of a stimulus as a discriminative stimulus was necessary as well as sufficient for secondary reinforcement. Guttman and Verplanck[2] followed this line of reasoning in a slightly different direction and suggested that primary as well as secondary reinforcement may derive fr the same principle. Their suggestion implies that primary reinforcing stimuli are effective because they are strong discriminative stimuli for a subsequent response which is sometimes called the consummatory response. The effects of deprivation on learning would be interpreted in terms of its effect on the strength of consummatory reflexes. The implications of this idea are far-reaching. However, further elaboration is beyond the scope of the present paper.

For our purposes it will be necessary to make a more explicit statement of the notion that discriminative stimuli exhibit secondary reinforcing properties. We note that the stimulus is considered as having a double role: first, its effect on immediately subsequent behavior which we call discriminative stimulus value or stimulus strength, and second, its effect on preceding *S-R* bonds which we will call reinforcing value. The present suggestion is that these two functions are directly related. In order to make a quantitative postulate on the basis of these considerations, we introduce the following definitions:

R = some class of "active" responses characterized by the fact that any member removes the subject from one stimulus to another.

Reflex = an *S-R* pair. No unconditioned connection is implied.

L = the total time that the stimulus of a reflex is present before the response occurs. This variable is not the same as the latency, in all cases, since the interval L need not be a continuous interval. For example, if R_1 failed to occur the first time S_1 was presented but occurred the second time, the value of L would be taken as the sum of the two intervals, whereas the latency would be the second interval only. The use of this variable will be indicated below.

V = *strength of a reflex* $= 1/L$. This measure of reflex strength was selected partly because there exists a correspondence between it and *momentary probability* which has proved to be a convenient variable in theories such as Estes' statistical theory of learning (3). This correspondence will be discussed in more detail later.

V_t = *strength of a stimulus* = the strength of the reflex S-R_t, where R_t is the combined class of all active responses to a particular stimulus. V_t is called the strength of the stimulus because it refers to a particular stimulus but not to any particular response. It will be noted that $1/V_t$ is equal to the latency.

We are now in a position to present the quantitative rule of secondary reinforcement to be used in the present model. We postulate that if a response R_1 removes a subject from S_1 to S_2,

[2] GUTTMAN, N., & VERPLANCK, W. S. Personal communication.

the strength of the reflex S_1-R_1 is changed as follows:

$$\text{delta } V_1 = c(V_{t2} - V_1) \qquad \text{[Rule 1]}$$

where

$V_1 =$ the strength of the reflex S_1-R_1,
$V_{t2} =$ the strength of the stimulus S_2, and
$c =$ a constant.

The equation states that V_1 is changed in the direction of V_{t2} by an amount proportional to the difference between them. If V_{t2} is greater than V_1, an increase in V_1 occurs; if it is less than V_1, a decrease occurs.

It will be seen that this rule is consistent with the previous qualitative statement of secondary reinforcement. In addition, it can be interpreted to include primary reinforcement if we assume the existence of at least one stimulus with a fixed maximum strength which will represent the primary reinforcing stimulus. If we insert a constant in place of V_{t2} in the above equation we obtain a function identical to the linear operator used by Bush and Mosteller (1). If we treat this equation as a differential equation we obtain functions formally identical to functions used by several theorists to describe simple learning (7, 3, 1).

In addition to the above rule of reinforcement we have found that it is necessary to the internal consistency of the model to include a second rule which relates to responses which fail to occur. The rule adopted states that if a stimulus S_1 is presented and the response R_1 fails to occur before some other response removes the subject from S_1 the reflex S_1-R_1 is changed as follows:

$$\text{delta } V_1 = -cV_1. \qquad \text{[Rule 2]}$$

This equation states that V_1 is decreased by a constant proportion of its value. The reason for including this rule will be discussed in more detail later.

These two rules are the only rules of change of reflex strength necessary to the present theory if we deal with "one-way" situations; that is to say, situations in which the subject never returns to the same stimulus twice during an experimental trial. An additional rule, which we will not elaborate, is required if retracing is allowed. Briefly, this rule provides for a decrement in any change in reflex strength if the subject returns to a stimulus soon after leaving it.

These postulates can be used to account for a variety of experimental findings in the sense that they yield behavior which is at least qualitatively the same as would be expected on the basis of experimental findings. This correspondence holds for findings of experiments on simple conditioning, extinction, delay of reinforcement, secondary reinforcement, and simple discrimination learning. The model also yields learning of observing responses under certain conditions and therefore can account for some kinds of "learning set," "concept formation," and stimulus generalization (11). It is expected that the model will also yield plausible predictions for intermittent reinforcement experiments and at least one kind of latent learning, although these cases have not been tested. The description of the electronic model and its operation will illustrate the way in which these postulates operate in the case of observing response learning in a two-choice situation.

THE ELECTRONIC MODEL

The electronic model consists of a group of variable time delay circuits, each representing a reflex. Each unit activates a relay when an input representing its stimulus has been connected for a cumulative total of L seconds.

The relay disconnects its own input circuit and connects the input for some other unit or units. The connections between the response of one unit and the stimulus of another are arranged by the operator so that all units and their connections form a *net* which represents some experimental situation. For example, a simple T maze might be represented by a net as follows:

$$\begin{array}{ll} & \text{-}R_LS_L \\ S_1\text{-}R_1S_2 & . \\ & \text{-}R_RS_R \end{array}$$

This net is constructed of three reflex units S_1-R_1, S_2-R_L, and S_2-R_R. If the experimenter presented S_1, the S_1-R_1 time interval would begin. After L_1 seconds, R_1 would occur, removing S_1 and connecting S_2. At this time, both the S_2-R_L and S_2-R_R units would begin timing. If R_L occurred first, it would remove S_2 and connect S_L. Thus the machine's "choice" is determined by the time intervals of the units. On the next trial when S_2 was presented the two S_2 units would continue timing. Note that the S_2-R_R unit will already have some time registered from the preceding trial since R_R did not occur. Thus this unit will have a "head start." In this description we have ignored changes in reflex strength and simply illustrated the way in which the responses of the device are controlled by the delay circuits. We will now consider changes in reflex strength.

According to the mathematical model, changes in the strength of a reflex are determined by (*a*) the occurrence or nonoccurrence of the response before the stimulus is removed by another response, and (*b*) the strength of the following stimulus. The strength of the following stimulus is reflected in the latency of the following response, which we have noted is the inverse of the stimulus strength. For example, in the net described above changes in V_1 are to be determined by the latency (L_{t2}) of the subsequent response to S_2. If the response to S_2 is sufficiently prompt, V_1 is to be increased; otherwise it is to be decreased. The amount of change prescribed by Rule 1 was given as $c(V_{t2} - V_1)$ which is equivalent to $c(1/L_{t2} - V_1)$. For reasons of technical expediency, a negative growth function was substituted in the machine for the reciprocal function. This substitution provides a fair approximation over the critical range and greatly simplifies the circuit design.

If a response fails to occur before the stimulus is removed by the occurrence of some other response, V_1 is to be decreased by a constant proportion of its value. These changes in reflex strength and the corresponding changes in time delays are produced automatically by the electrical circuits. A value of .5 was used for the constant c in the present model. To save space, circuit diagrams and technical details will be omitted from the present report.[3] We are primarily concerned with the learning theory at this time. However, we may mention that the device employs relatively simple thyratron timer circuits. The strength of each reflex is represented by a voltage on a storage condenser, and this voltage is changed according to the postulates of the theory by appropriate relay connections.

Note that the strength of a reflex is affected by events which follow it in time, so that a unit which has just operated must be registered until the next

[3] To save printing costs, the circuit diagrams for the present model together with an eight-page discussion of the circuit's operation have been deposited with the American Documentation Institute. Order Document No. 4160 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance \$1.75 for 35 mm. microfilm or \$2.50 for 6 by 8 in. photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

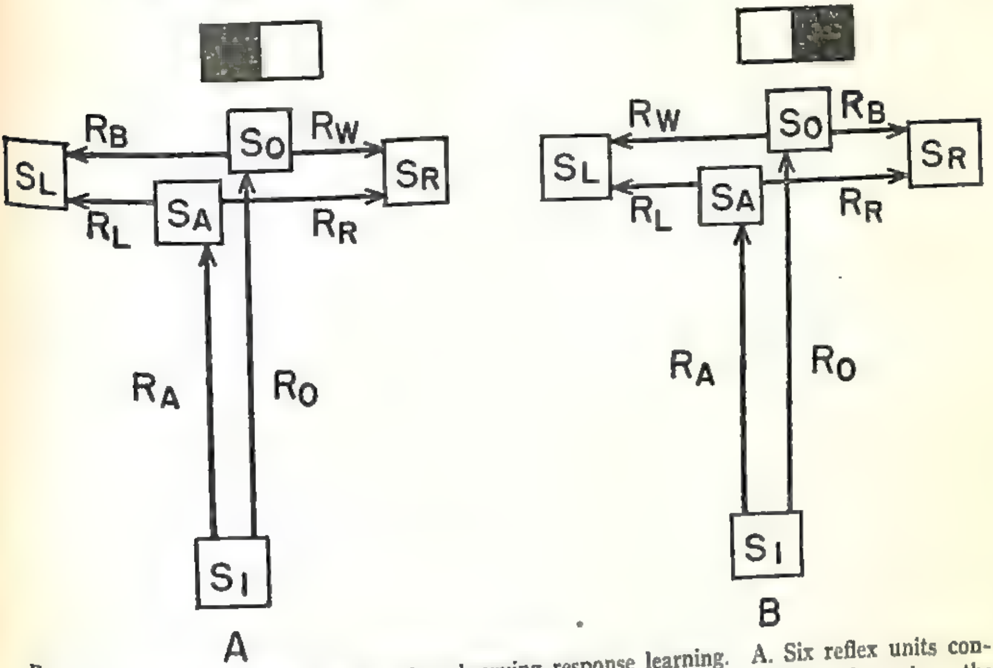

FIG. 1. The reflex net used to test for observing response learning. A. Six reflex units connected to represent a T maze with two stimulus cards at the choice point; a black card on the left and a white card on the right. B. The same T maze with the stimulus cards reversed.

response occurs. This system represents a compromise between a system such as Estes', which requires no "traces" or delayed effects, and other systems which require prolonged traces, or effects dependent on the final outcome of a trial.

All reflex units are identical in structure with the exception of a "primary reinforcing unit." This unit is a fixed interval timer with an interval of approximately .2 sec. The intervals of the other units may vary between approximately .2 and 80.0 sec. The units may be connected in any arbitrary net and may be attached to an array of lights on a panel representing some experimental situation. The operator may "run experiments" in much the same way that a rat experiment would be run, "placing the subject" at a starting point and connecting the primary reinforcing unit to some arbitrary position in the net. Latencies, rates, or relative frequencies of various responses can be recorded. In the following, we will describe a particular net which was set up to test for the learning of observing responses in a two-choice situation. This net is represented in Fig. 1. The diagram represents a T maze with two stimulus panels at the choice point. The experimenter may light either one so that it appears white while the other appears black.

From the starting box (S_1), two responses may occur, R_a or R_o. R_a may be thought of as the response of approaching the choice point with head down so that the stimulus panels are not visible. From S_a the subject may turn right (R_R) or left (R_L), but these responses will be independent of the position of the black and white panels. R_o, on the other hand, may be thought of as approaching the choice point with head up, so that the panels are visible. Responses from S_o are completely dependent on the position of the stimulus panels. They are specified as approaching black (R_B) and approaching white (R_W). R_W takes the subject to the right goal box if the right panel is lighted and to the left goal box if the left panel is lighted. R_B is the opposite response.

The machine was connected to an array of lights representing the various positions in the net and the stimulus panels. The operator was then able to run experiments by switching on one of the stimulus lights, connecting the reinforcement unit to one of the goal box terminals, and presenting S_1. At this time, the S_1 panel light would appear. After an interval the light would "jump" to S_a or S_o, and then to one of the goal boxes. If it arrived at the correct goal box, the reinforcement unit would operate, turning off the goal box light after about .2 sec. If it arrived at the wrong goal box, the goal box light would simply stay on until the experimenter started the next trial.

Note that the reinforcement unit operates any time the subject arrives at the correct goal box, regardless of whether the initial response was R_a or R_o. Thus the subject may be reinforced even though it does not make the correct observing response. This corresponds to the fact that a subject in a discrimination experiment may make the correct turn even though it is not exposed to the discriminative stimuli. The experimenter places food in one of the goal boxes and the subject obtains the reinforcement when it arrives at the correct goal box even if it runs with its eyes closed. This general characteristic of situations involving observing responses is, in part, the source of the difficulties of predicting observing response learning from a quantitative theory.

Four kinds of discrimination problems can be tested in this net. In two of them "color" will be relevant, reinforcement always being presented for running toward one of the two colors. In the other two, "side" will be relevant, reinforcement always being presented for running to a particular side. Note that color discriminations can be solved only if the initial response is predominately R_o. Side discriminations depend on an initial response of R_a. We assume that the stimulus lights are reversed at random. Our interest is focused largely on whether the machine will learn to make R_a or R_o depending on whether reinforcement is consistently on one side, or consistently on one color. We will examine the results of one experiment in which the machine was confronted with six successive discrimination problems. The sequence of problems described in terms of the positive stimulus was as follows: white, black, white, right, left, black. Thus in the first problem, reinforcement was always connected to the goal box opposite the lighted stimulus panel regardless of which side this might be. In the fourth problem the reinforcement was always connected to the right-hand goal box, etc. This sequence provides several illustrations of reversals (white to black, right to left, etc.) and two illustrations of changes in the relevant aspect (white to right and left to black).

RESULTS

Figure 2 presents a learning curve for the entire sequence of six discrimination problems. The solid lines represent percentage of correct choices, that is, the percentage of response sequences terminating in the correct goal box. The dotted lines represent the percentage of correct observing responses. R_o was considered correct during color discriminations, R_a during side discriminations.

The curves in the first section of the graph show that both the correct choice and the correct observing response were learned while "white" was the positive

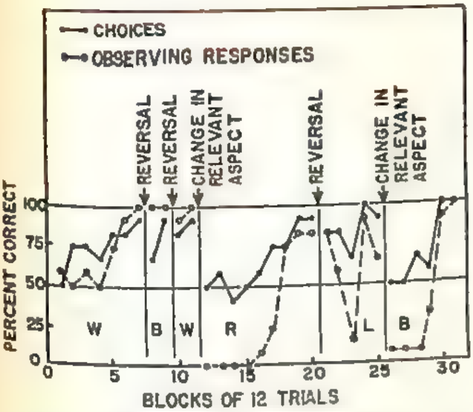

Fig. 2. Performance of the electronic model on six successive discrimination problems. The letters W, B, R, and L indicate the positive stimulus for each problem.

stimulus. In the second and third sections we note that reversed discriminations are learned very rapidly. The observing response which was learned during the first problem was still correct after the reversals and the percentage of correct observing responses remained high. On the fourth problem the relevant aspect was changed from color to side. The observing response which had been correct during the previous problems was now incorrect. We notice considerable retardation in learning on this problem. Learning of the correct observing response did not begin until after the fourth block of 12 trials. However, once this learning was started it proceeded fairly rapidly, reaching a reasonably high level within the next three blocks of 12 trials. In the fifth problem, following the reversal from right to left, we notice a temporary drop in the observing response curve, even though the same observing response was correct. This drop may be attributed to chance contingencies between incorrect observing responses and correct choices. Such contingencies may temporarily strengthen the incorrect observing response. Thus the learning of the reversed discrimination was also retarded in this case. The sixth problem shows the second change in relevant aspect and again shows retardation of learning at the beginning, followed by relatively rapid learning of the correct responses. In this case, the break in the curve was so sharp that it gave the impression of "insight" learning.

Discussion

The present findings serve to illustrate the way in which a physical model may facilitate the development of a theory. By its use it was possible to show that the postulates of the theory predict the learning of observing responses and that, in this case, nothing outside the realm of this theory is necessary to account for observing response learning. Critics of reinforcement theory have sometimes argued that the facts of "attention" are incompatible with *S-R* reinforcement theories and that these facts compel us to make a basic revision in our theorizing. Hebb (6), for example, uses this argument to justify the introduction of a "central autonomous process." The present findings suggest that such critics underestimate the potential explanatory power of reinforcement theory.

The present development has also provided a clear illustration of the way in which a mechanical model may reveal the shortcomings of a theory. We mentioned earlier that the initial model failed to show observing response learning. A serious error in the original theory was not evident until the machine forced it to our attention. The original model operated on the basis of Rule 1 alone, so that no change in reflex strength was made for responses which failed to occur. When this model was tested on the observing response net (Fig. 1), it consistently learned not to make the correct observing response. The source of the difficulty, briefly stated, was that the strength of S_a remained too high during color discrimination learning. The combined strength of R_R and R_L proved to be greater than the combined strength of R_W and R_B. On the basis of this observation the theory was re-examined. It became apparent that some decrement in the strength of responses which failed to occur would be necessary. It was possible to prove this necessity entirely apart from the problem of observing response learning. Rule 2 was added to correct this discrepancy, and further tests showed that the corrected model would yield observing response learning.

The variable V. The present model makes use of a somewhat novel measure of reflex strength, the variable V. The use of this variable enables us to maintain a certain degree of correspondence with a statistical model such as those proposed by Estes (3), Bush and Mosteller (1), and Estes and Burke (4), and at the same time allows for relatively simple mechanization of the model. The correspondence to the probability model will be evident if we examine the relationship between V and other measures of reflex strength. First, in a single-choice situation the reciprocal of V is the latency. In a two-choice situation, where two responses have strengths V_1 and V_2 the average latency will be $1/(V_1 + V_2)$. The relative frequency of R_1 will be $V_1/(V_1 + V_2)$, and of R_2 will be $V_2/(V_1 + V_2)$. In a free responding situation such as a Skinner box, the rate will be proportional to V. Parallel relationships hold if we use momentary probability (p) as a measure of reflex strength, where p is defined as the probability of occurrence of a particular response in a short time interval h. If we assume that p is the same at any moment during a given trial, the quantity p/h may be substituted in the above functions to obtain corresponding values for mean latencies and relative frequencies. (To obtain this correspondence it is necessary to assume that the time interval h is chosen sufficiently small so that the probability of two different responses occurring in the same interval is negligible.) It should be made clear that this correspondence holds for mean values only. The models do not correspond with respect to the distributions of these variables. This difference may or may not prove to be critical.

It is interesting to note that the present model, which was originally based on propositions growing out of reinforcement theory, shows some marked similarities to a contiguity theory, in that reinforcement is closely associated with removal of stimuli. However, if we examine the situation more closely we see that different stimuli are involved in the two cases. In Guthrie's system (5) the critical factor is the removal of the stimulus which was present at or before the time the response occurred, whereas in the present model reinforcement depends on removal of the stimulus which appears after the response is made, and we require that the stimulus be removed by the occurrence of a subsequent response.

SUMMARY

The present mathematical model was formulated with the objective of developing a quantitative model which would take into account the learning of observing responses. A quantitative postulate of secondary reinforcement plays an important role in this formulation. The postulate is based on a qualitatively expressed notion suggested by Skinner that secondary reinforcing properties of a stimulus are related to the discriminative stimulus value of the stimulus. By adopting quantitative definitions of reinforcement and "discriminative stimulus value" this proposition was readily translated into a quan-

titative postulate. This postulate forms the core of the mathematical model.

An electronic model was constructed to test whether the postulate system would yield plausible predictions of observing response learning. The device constitutes a robot which operates according to the postulates of the mathematical model. The electronic model was confronted with a discrimination problem in which it was required to select, by means of observing responses, those aspects of the stimulus situation which were relevant. The operation of the machine demonstrates that the theory will yield observing response learning. Some of the implications of the theory and the use of a physical model in theory construction are discussed.

REFERENCES

1. Bush, R. R., & Mosteller, F. A. A mathematical model for learning. *Psychol. Rev.*, 1951, **58**, 313–323.
2. Dinsmoor, J. A. A quantitative comparison of the discriminative and reinforcing functions of a stimulus. *J. exp. Psychol.*, 1950, **40**, 458–472.
3. Estes, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.
4. Estes, W. K., & Burke, C. J. A theory of stimulus variability in learning. *Psychol. Rev.*, 1953, **60**, 276–286.
5. Guthrie, E. R. *The psychology of learning.* New York: Harper, 1935.
6. Hebb, D. O. *The organization of behavior.* New York: Wiley, 1949.
7. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
8. Notterman, J. M. A study of some relations among aperiodic reinforcement, discrimination training, and secondary reinforcement. *J. exp. Psychol.*, 1951, **41**, 161–169.
9. Schoenfeld, W. N., Antonitis, J. J., & Bersh, P. J. A preliminary study of training conditions necessary for secondary reinforcement. *J. exp. Psychol.*, 1950, **40**, 40–45.
10. Skinner, B. F. *The behavior of organisms.* New York: Appleton-Century, 1938.
11. Wyckoff, L. B. The role of observing responses in discrimination learning: Part I. *Psychol. Rev.*, 1952, **59**, 431–442.

(Received April 1, 1953)

Psychological Review
Vol. 61, No. 2, 1954

A STATISTICAL THEORY OF THE PHENOMENON OF SUBCEPTION

DAVIS HOWES[1]

Aero Medical Laboratory, Wright-Patterson AFB, Ohio

Recent evidence reported by Lazarus and McCleary (3) leads to the conclusion that a conditioned autonomic response is a more sensitive indicator of the recognition of nonsense syllables than is the observer's symbolic report. For this phenomenon—sometimes described as autonomic discrimination without awareness—they have suggested the term *subception*. Its existence is of special interest because it invalidates a generalization that appears to hold for a great variety of perceptual experiments. This generalization may be formulated as follows: when an observer is given the task of discriminating among a set of stimuli, no measure of his success at that task is more sensitive than his symbolic (verbal) report. We shall refer to this as the *symbolic-report hypothesis*. Previous attempts to invalidate it have been numerous, but, as Lazarus and McCleary show in their report (3, pp. 113–116), satisfactory experimental controls were not observed in these earlier studies. Their own data, obtained in a very carefully designed experiment, thus provide the principal evidence against the hypothesis. It is the purpose of this paper to show that these data actually are not inconsistent with the symbolic-report hypothesis when certain statistical effects are taken into consideration.

THE EXPERIMENT

The Lazarus and McCleary experiment consisted of two essential parts, a conditioning period and a test period. During the conditioning period, galvanic skin responses (GSR), elicited by electric shock, were conditioned to each of 5 nonsense syllables (shock syllables) exposed in a tachistoscope for 1 sec. Five other nonsense syllables (nonshock syllables) were exposed an equal number of times, but were not paired with shock. Observers were shown the list of 10 syllables and were told which 5 would be paired with shock and which 5 would not. In the test period, the 10 syllables were again exposed, but at 5 different durations, selected by pretest to range from durations too brief for accurate recognition to durations long enough to result in almost 100 per cent recognition. No shocks were given during the test period. The important data recorded during the test period were the GSR for a 5-sec. period following each exposure, and the observer's verbal report as to which of the 10 syllables had been exposed. Precautions taken by *E*s to control various artifacts are summarized in their report (3, pp. 116–118).

Three properties of the observer's reports were used to categorize the data. These, with the symbols used to represent them, are as follows: (*a*) the report was right (*R*) or wrong (*W*); (*b*) the report was a shock syllable (*S*) or a nonshock syllable (*N*); (*c*) the stimulus actually exposed was a shock syllable (subscript *s*) or a nonshock syllable (subscript *n*). The possible combinations of these properties define the six categories RS_s, WS_s, WN_s, RN_n, WS_n, WN_n. For each of these experimental categories Lazarus and McCleary report GSR

[1] Present address Wilmington, N. C.

in micromhos, averaged over all five exposure times.

A distinction needs to be drawn here between the *subception effect* and the *subception hypothesis.* The subception effect is defined by Lazarus and McCleary as the inequality

$$(WS_s + WN_s) > (WS_n + WN_n),$$

where the symbol for each category represents the mean GSR obtained under the conditions defining that category. This inequality was found to hold for all nine observers used in the Lazarus and McCleary experiment, and the mean difference in GSR could not be accounted for by chance variations ($t = 7.45$, $df = 8$). These results establish the existence of the *subception effect.* This effect leads Lazarus and McCleary to conclude (p. 118) that "subjects can make autonomic discriminations when they are unable to report conscious recognition" (i.e., correct symbolic report). Such a mechanism of "unconscious perception" can be called the *subception process.* No property of the experimental data other than the subception effect is offered by Lazarus and McCleary as evidence for a process of this type. It will be argued here that no subception process need be postulated, since the subception effect can be derived from the symbolic-report hypothesis.

The Statistical Theory

Let us take for a starting point a statement of the symbolic-report hypothesis that is simply a negation of the subception hypothesis stated by Lazarus and McCleary: *GSR discrimination can occur only when the observer can report the exposed syllable correctly.* This formulation, which we shall refer to as the *classical* form of the symbolic-report hypothesis, is invalidated by the existence of the subception effect. But it does not take into account the statistical nature of discrimination. The theory proposed here considers the processes underlying recognition to be stochastic variables. This conception leads to the *statistical* form of the symbolic-report hypothesis: *at any specified moment, the GSR accompanying an observer's report is proportional to the probability that that report will be a shock syllable.* Like its classical counterpart, this formulation of the hypothesis implies that there is no autonomic discrimination without verbal discrimination, but it treats verbal report as a statistical process. A similar assumption has been proposed elsewhere (1) to account for the unusually long exposure times found to be required for the recognition of taboo words (4).

From the statistical form of the hypothesis it follows that any condition which alters the probability that a shock syllable will be reported will elicit a corresponding change in GSR. Tachistoscopic exposure of a shock syllable is such a condition. We shall follow the argument for a set of m syllables so exposed that they can be read when exposed for indefinitely long durations. Consider the effect of exposing a syllable for the limiting conditions of zero and of infinite duration. For zero duration of exposure all syllables will have the same probability of report if (as in the Lazarus and McCleary study) the m syllables are distributed equally over the experimental conditions. The probability of any exposed syllable is thus $1/m$. When the syllable is exposed for indefinitely long durations, however, the probability that it will be reported correctly approaches 1. These limiting conditions show that increasing the exposure time of a given syllable from zero to infinite duration raises the probability that the syllable will be reported. All experimental evidence indicates that between these limiting conditions the relation between probability of correct verbal report and exposure duration is monotonic and nondecreasing (2, 5, 6, 7). We then have the implication

$$[t_2(i) > t_1(i)] \subset [p_2(i) \geq p_1(i)], \qquad [1]$$

where $p_2(i)$ is the probability that syllable i will be reported following its exposure for $t_2(i)$ seconds and $p_1(i)$ the probability of its report following exposure for $t_1(i)$ seconds. The symbolic-report hypothesis, in the statistical form from which the subception effect is to be derived, may be written

$$[p_2(i_s) \geq p_1(i_s)] \subset [c_2(i_s) \geq c_1(i_s)], \qquad [2]$$

where i_s represents a *shock* syllable, $c_2(i_s)$ is the mean palmar conductance (GSR) measured when $p_2(i_s)$ is the probability of the observer's reporting i_s, and $c_1(i_s)$ is the mean GSR when $p_1(i_s)$ is the probability of his reporting i_s. From Equations 1 and 2 we have, by *modus ponens*,

$$[t_2(i_s) > t_1(i_s)] \subset [c_2(i_s) \geq c_1(i_s)]. \qquad [3]$$

This conclusion states that, on the average, exposure of a shock syllable for a long duration will yield a larger GSR than exposure of the same syllable for a shorter duration.

Now of the two pairs of experimental categories that define the subception effect, $(WS_s + WN_s)$ includes only shock-syllable exposures while $(WS_n + WN_n)$ includes only exposures of nonshock syllables. Thus $t(i_s)$, the time for which shock syllables are exposed, must be zero for $(WS_n + WN_n)$ and greater than zero for $(WS_s + WN_s)$; and the inequality of GSR that defines the subception effect,

$$(WS_s + WN_s) > (WS_n + WN_n), \qquad [4]$$

follows from Equation 3. Our argument to this point shows that the existence of the subception effect does not require that the symbolic-report hypothesis be abandoned, but only that the formulation of the hypothesis take into account the fact that recognition must be defined in statistical concepts.

FURTHER DEDUCTIONS FROM THE STATISTICAL THEORY

We can deduce more than the gross subception effect [4] from our statistical hypothesis. Lazarus and McCleary report mean GSR for all six of their experimental categories. A deduction can be drawn from our assumptions for each of the 15 possible comparisons between these means. To make the argument clear, however, we must analyze the concept of response probability a little more closely.

Corresponding to each of our m syllables we assume m characteristic underlying processes, each independent of the others. These processes will be regarded as analytic fictions, and no physiological referent need be ascribed to them. An observer will emit (report) the ith syllable only if the strength of the ith process is greater than the strength of the process underlying any of the other syllables. At this point we introduce our stochastic variable: we assume that the strength of the ith process, w_i, is determined stochastically, not uniquely, by the specification of experimental conditions. While we cannot fix w_i by experimental techniques, we can use such techniques to fix a distribution function $F(w_i)$. Suppose a set of conditions fixes the same distribution function for two syllables i and j. If the momentary strengths of the two syllables are observed upon an infinitely large number of occasions,

w_i will be the greater on half of these observations, w_j on the other half. Since the emission of a syllable may be considered to depend upon the integration of these momentary strengths over short periods of time, the probability (limit of relative frequency) of emission of syllable i will be equal to that of syllable j. We can therefore associate the condition of identical distribution functions with syllable probabilities of $1/m$ and with zero duration of exposure. The other limiting condition, where all values of the distribution function of the ith process are greater than the largest possible value of the process underlying any other syllable, will yield a probability of i equal to 1 and may therefore be associated with the condition of infinite exposure of i.

All we have done in the above argument is to identify an increase in the duration of exposure of the ith syllable with an increase in the probability that the ith process will exhibit greater strength than the process underlying any other syllable. The symbolic-report hypothesis can now be reformulated: at any specified moment, *the GSR accompanying an observer's report is proportional to the total strength of all processes underlying shock syllables.*

We must now assess the relative strengths of the shock-syllable processes in each of the six categories of the Lazarus and McCleary experiment. It must be remembered, however, that the report of the observer only tells us which of the various underlying processes is the strongest. Yet a shock syllable may contribute heavily to the GSR even if it is not reported by the observer, for its underlying process may be only slightly weaker than the strongest process. How can the relative strengths of these unreported syllables be analyzed? For the data reported by Lazarus and McCleary there is no way to obtain this information for a given exposure, but the assumptions we have made will suffice to determine averages. We shall restrict our analysis to the strongest and second-strongest underlying processes, since an analysis of third-strongest and still weaker processes would simply repeat the argument and reinforce the effects shown for the second-strongest processes.

Before proceeding to the argument, it will be convenient to summarize here the definitions and assumptions upon which the argument will be premised. *Definition:* Verbal report of the ith syllable means that, at the moment of report, the strength of the ith process is greater than the strength of any other process. *Assumption 1:* The probability that the ith process will have a strength w_i is a monotonic, nondecreasing function of the duration for which syllable i is exposed. *Assumption 2:* Exposure of any syllable other than the ith will have only a negligible effect upon the probability that the ith process will have a strength w_i. *Assumption 3* is the symbolic-report hypothesis. The argument will be carried through only for the special case in which all unexposed syllables are assumed to have identical distributions of process strength (i.e., are equally likely to be reported), although it can be extended to cover the general case in which each unexposed syllable is assigned a different distribution of strength.

Let the capital letters A through E represent the processes underlying the five shock syllables A' through E' of the Lazarus and McCleary experiment, and let the lower-case letters a through e represent the processes underlying the five nonshock syllables a' through e'. Consider the situation when a shock syllable, A', is exposed.

All observer reports placed in the RS_s category must then have A for their strongest underlying process. But by our assumptions, no one of the remaining nine processes is more likely than any other to be the second-strongest process. Consequently, nine equally probable combinations of strongest and second-strongest processes are possible for this category: (A, B), (A, C), (A, D), (A, E), (A, a), (A, b), (A, c), (A, d), (A, e). These alternatives are listed under RS_s in Table 1.

TABLE 1

POSSIBLE COMBINATIONS OF STRONGEST AND SECOND-STRONGEST SYLLABLE PROCESSES FOR EACH OF THE SIX CATEGORIES OF THE LAZARUS AND MCCLEARY EXPERIMENT

A' Exposed	a' Exposed
RS_s	RN_n
A A A A A A A A A B C D E a b c d e	a a a a a a a a a A B C D E b c d e
WS_s	WS_n
B B B B B B B B B A C D E a b c d e	A A A A A A A A A B C D E a b c d e
C C C C C C C C C A B D E a b c d e	B B B B B B B B B A C D E a b c d e
D D D D D D D D D A B C E a b c d e	C C C C C C C C C A B D E a b c d e
E E E E E E E E E A B C D a b c d e	D D D D D D D D D A B C E a b c d e
	E E E E E E E E E A B C D a b c d e
WN_s	WN_n
a a a a a a a a a A B C D E b c d e	b b b b b b b b b A B C D E a c d e
b b b b b b b b b A B C D E a c d e	c c c c c c c c c A B C D E a b d e
c c c c c c c c c A B C D E a b d e	d d d d d d d d d A B C D E a b c e
d d d d d d d d d A B C D E a b c e	e e e e e e e e e A B C D E a b c d
e e e e e e e e e A B C D E a b c d	

Still assuming that A' is the syllable exposed, let us examine the WS_s category. Here the strongest process must by definition correspond to one of the unexposed shock syllables B', C', D', or E', and by our assumptions there is an equal probability that it will be each one. But the second-strongest process cannot be specified so simply. Take the case where B is the strongest process. Any of the nine remaining processes can be the second-strongest. But one of these is A, the process underlying the exposed syllable, and by our first assumption its average strength depends upon the duration for which A' has been exposed. Only for exposures approaching zero duration will the average strength of A be as small as the average strengths of the other eight processes. When A' is exposed and B' happens to be reported, then, the second-strongest process will probably be A. The remaining eight processes will have smaller probabilities of being second-strongest. The difference between the probability of A and the probability of each of the other processes will depend upon the duration for which A' has been exposed: for short durations the difference will be small; for long durations the probability that A is the second-strongest process will approach unity.

We have analyzed the WS_s category only for the case in which B' is the syllable actually reported. The same analysis obtains for the equally likely cases where C', D', or E' happens to be reported. In each case, A is the most likely process to be second-strongest. The 36 possible combinations of strongest and second-strongest processes for this category have been listed in Table 1. The combinations (B, A), (C, A), (D, A), and (E, A), which are more probable than the others whenever A' has been exposed, have been printed in boldface.

Each of the other four categories used by Lazarus and McCleary has been analyzed in the same way, and all possible combinations of their strongest and second-strongest processes are listed in Table 1. The table is worked out for the cases where A' is the shock syllable exposed and a' is the nonshock syllable exposed. Combinations printed in the same typeface are of equal probability; those printed in boldface are the combinations whose probabilities are increased by exposure of A' or a'. For exposure durations approaching zero the probabilities of the boldface combinations approach those of the other combinations; for long exposures the probabilities of the boldface combinations become so great that the other combinations may be ignored.

Table 1 permits us to compare the shock-syllable strengths of any two categories for specified exposure durations. Now Lazarus and McCleary report only *average* GSR for all reports falling within each experimental category. In order to compare shock-syllable strengths averaged over all exposure durations, then, we must consider the extent to which each category will contain reports following long and short exposures. For exposures of zero duration, the probability that the "exposed" syllable will be reported will be no greater than the probability that any other syllable will be reported. The probability that a report will fall into the RS_s category if a shock syllable is exposed (or into the RN_n category if a nonshock syllable is exposed) will thus be only 1/10. For long exposures, however, the probability of the exposed syllable will be high, and all but a few of the observer's reports will fall into the RS_s (or RN_n) category. Most of the reports falling into the RS_s and RN_n categories will thus have been emitted following long exposures, while most of the reports falling into the other four categories will have been emitted following short exposures. This heterogeneity must be taken into account in comparing either RS_s or RN_n with any of the other categories.

We are now ready to determine which of each pair of categories will have the greater shock-syllable strength. In 8 of the 15 possible pairs of categories the strongest process of one category underlies a shock syllable while the strongest process for the other category underlies a nonshock syllable, and the proportion of long and short exposures in the two categories either will not differ appreciably or will reinforce the difference between their strongest processes. Inequalities of shock-syllable strength that will hold regardless of exposure duration can therefore be stated for these pairs of categories. Hence $RS_s > RN_n$; $RS_s > WN_s$; $RS_s > WN_n$; $WS_s > RN_n$; $WS_s > WN_s$; $WS_s > WN_n$; $WS_n > RN_n$; $WS_n > WN_n$. In these inequalities the symbol for each category represents the shock-syllable strength averaged over all responses falling into that category.

Table 1 contains another four pairs for which one category has greater shock-syllable strength for all durations of exposure except those approaching zero. At durations approaching zero the shock-syllable strengths of the compared categories approach equality. These also are pairs for which the proportion of long and short exposures either will not differ appreciably or will reinforce the difference apparent from Table 1. Since shock-syllable strengths must be averaged over all exposure durations, inequalities may also be stated for these pairs: $WS_s > WS_n$; $WN_s > WN_n$; $RS_s > WS_n$; $WN_s > RN_n$. The first two of these, it may be noticed, correspond to the subception effect reported by Lazarus and McCleary.

Consider next the RS_s and WS_s categories. Table 1 indicates that both categories will have equal shock-syllable strengths for exposure durations approaching zero. But for very long exposures almost all of the second-strongest processes in the WS_s category will underlie shock syllables, while 4/9 of the second-strongest processes in the RS_s category will underlie shock syllables. Reports following exposures of a specified long duration, then, should reflect a greater average shock-syllable strength if they fall into the WS_s category rather than the RS_s category, though the difference, being restricted to a portion of the second-strongest processes in the respective categories, will be small. The RS_s category, however, will contain up to 100 per cent of the reports following long exposures, while the WS_s category will include up to 4/5 of the short exposures. When shock-syllable strengths are averaged over all reports included in each category, then, the inequality will be $RS_s > WS_s$, for RS_s will be composed chiefly of large strengths derived from long exposures while WS_s will be composed chiefly of small strengths derived from short exposures.

The RN_n and WN_n categories also are identical for very brief exposures. Increasing exposure duration will not alter the shock-syllable strength of RN_n. The shock-syllable strength of WN_n, however, will *decrease* with increasing exposure duration, because the boldface combinations for this category include nonshock syllables as both strongest and second-strongest processes. On the average, therefore, reports following long exposures will reflect slightly smaller shock-syllable strengths if they fall into the WN_n category than if they fall into RN_n. But practically all the long exposures will fall into the RN_n category, while up to 4/5 of the short exposures will fall into the WN_n category. Pooling exposures of all durations will therefore reduce the average difference between these categories to a very small value.

Comparison of WS_n and WN_s involves a similar problem. The inequality $WS_n > WN_s$ will hold for short exposures. But increasing the exposure time will increase the average strength of the second-strongest processes of the boldface combinations shown for these categories in Table 1. As a consequence, the average difference between the strongest and second-strongest processes of each of these categories will decrease with increasing exposure, since the strongest process in each will underlie an unexposed syllable and hence will not gain appreciably in strength as a result of the exposure. With exposures for which nearly 100 per cent of reports are correct, the difference between the shock-syllable strengths of these categories will be negligible. These long exposures, for which the two categories approach equality, will produce greatest shock-syllable strengths and will therefore have the greatest weight in determining the average shock-syllable strength for each category.

In the last two comparisons, between RN_n and WN_n and between WN_s and WS_n, inequalities of shock-syllable strength can be shown only for exposure durations which produce small process strengths. For exposure durations producing maximum process strengths, however, the shock-syllable strengths of each of these pairs of categories will approach equality. Since maximum process strengths will have the greatest weight in fixing the average shock-syllable strength of each category, the inequalities for these two pairs of categories will be seriously weakened. Except when the averages are taken over very large numbers of observations, these weak inequalities probably will be masked by the random variability that is a defined property of the syllable process and by the errors of measurement that will result from the effects of uncontrolled factors upon GSR. Now Lazarus and McCleary report the following numbers of observations per observer for the four categories involved: 24 (RN_n), 14 (WN_n), 13(WS_n), and 15 (WN_s). As the inequalities cannot be expected to hold reliably for such small numbers of observations, the relative shock-syllable strengths of these categories will be written $RN_n \approx WN_n$ and $WS_n \approx WN_s$, where the direction of each unreliable inequality is preserved in the order in which each pair of categories is written.

TABLE 2

Tests of the Deduced Inequalities

Column A lists the inequalities of shock-syllable strength. Column B shows the number of observers in the Lazarus and McCleary experiment (3) whose mean GSR confirms each inequality and the number whose mean GSR contradicts it.

The probability of obtaining 9 confirmations by chance is .002; 8 confirmations, .018; 7 confirmations, .070.

A	B	
	Yes	No
$RS_s > RN_n$	9	0
$RS_s > WN_s$	9	0
$RS_s > WN_n$	9	0
$WS_s > WN_s$	8	1
$WS_s > RN_n$	9	0
$WS_s > WN_n$	9	0
$WS_n > RN_n$	7	2
$WS_n > WN_n$	7	2
$WS_s > WS_n$	9	0
$WN_s > WN_n$	9	0
$RS_s > WS_n$	9	0
$WN_s > RN_n$	8	1
$RS_s > WS_s$	*7	1
	Total 109	7

* No difference in mean GSR could be detected between RS_s and WS_s for subject GW.

We have now analyzed the relative strengths of all processes underlying shock syllables for each of the 15 possible combinations of the experimental categories. Definite inequalities can be stated for 13 of these comparisons. According to the symbolic-report hypothesis, as it has been formulated in this section, the category for which the shock-syllable processes are stronger will have the greater GSR. This hypothesis can now be tested by comparing the 13 inequalities of shock-syllable strength with inequalities of GSR based upon the experimental data reported by Lazarus and McCleary. Table 2 presents this comparison. The left-hand column states the 13 inequalities derived above. The column on the right gives the number of observers in the Lazarus and McCleary study whose inequalities of GSR are in the predicted direction and the number in the opposed direction. The data are taken from their figure (p. 119). For none of the 13 inequalities do more than two of the nine observers show differences in GSR opposed to the predictions based on the statistical form of the symbolic-report hypothesis. Of the 116 inequalities that can be established from the Lazarus and McCleary data, only seven run counter to prediction.

Table 3 presents a similar tabulation for the two unreliable inequalities. These compare categories that cannot be expected to show reliable differences in GSR for the small numbers of observations that have been made. A chance distribution of GSR inequalities should therefore be found.

TABLE 3

TESTS OF THE TWO UNRELIABLE INEQUALITIES

(Arranged as in Table 2)

A	B	
	Yes	No
$RN_n \approx WN_n$	5	4
$WS_n \approx WN_s$	4	5
	Total 9	9

The data show just such a distribution: the nine inequalities of GSR for each of these pairs of categories are divided as equally as possible between those for which one category has the larger GSR and those for which the other category has the larger GSR. Of the 18 inequalities of GSR that can be set up from the experimental data, nine are in the same direction as the unreliable inequality of shock-syllable strength, and nine are in the opposed direction. Again the data are consonant with the analyses of shock-syllable strengths.

LOWER-ORDER REPORTS

Those assumptions about the recognition process that do not involve GSR can be tested directly by an experiment in which the observer reports more than one syllable following each exposure. Only the theory of these phenomena will be discussed here. The relevant data could be obtained from a modification of the recognition phase of the Lazarus and McCleary experiment in which the observer would be instructed to list, following each exposure, the m different syllables in the order in which they seemed to him most likely to have been the exposed syllable. Alternatively, the subject can be told that his initial report was incorrect and asked to give a different report. If this too is incorrect, he can be asked to give a third, etc., until he makes the correct report.

The initial or first-order report in either of these series can be identified with the single report usually recorded in a recognition experiment. We have defined the *strongest underlying syllable process* by this first-order report. Now suppose we eliminate this strongest syllable process from the alternatives available to the observer by asking him to give a different second-order report. The syllable then reported must, by the same definition, correspond to the *strongest remaining syllable process.* Thus the second-strongest syllable process will correspond to the second-order report. Similarly, if we eliminate $r - 1$ alternatives on the basis of the observer's reports, the rth-order report should correspond to the rth-strongest syllable process.

In the argument that follows we shall derive an expression for the probability that an exposed syllable will be emitted on the rth-order report if it has not been emitted on a previous report in terms of its first-order probability. We shall assume as before that a syllable process has a definite strength following each exposure and that this strength is a random variable over a series of exposures of the same duration. We shall also continue the simplifying assumption that the average strengths of all unexposed syllables differ only negligibly from one another, an assumption provided for by the counterbalanced design of the Lazarus and McCleary experiment.

Consider a matrix of the probabilities of emission for each of m possible syllables on r successive reports. Let i be the exposed syllable and α, β, γ, . . . be the $m - 1$ unexposed syllables. Table 4 shows such a matrix for $m = 6$. We wish to show that our assumptions determine

TABLE 4

Matrix Showing the rth-Order Probabilities of Emission for Each of Six Syllables

In this example it is assumed that $P_1(i) = .5$ and that the unexposed syllables $\alpha, \beta, \ldots, \epsilon$ are emitted on first-order, second-order, . . ., fifth-order report, respectively.

	r					
i	1	2	3	4	5	6
	5/10	5/9	5/8	5/7	5/6	5/5
α	1/10	0	0	0	0	0
β	1/10	1/9	0	0	0	0
γ	1/10	1/9	1/8	0	0	0
δ	1/10	1/9	1/8	1/7	0	0
ϵ	1/10	1/9	1/8	1/7	1/6	0

the entire matrix when only the first-order probability of the exposed syllable i is given.

Let us consider the first-order reports. Let $P_1(i)$ be the probability that the exposed syllable is emitted as the first-order report. Then the probability that one of the $m - 1$ unexposed syllables will occur as the first-order report is $Q_1(i) = 1 - P_1(i)$. The probability that any particular unexposed syllable, say α, will be the first-order report is $Q_1(i)/(m - 1)$, since the strengths of all unexposed syllables are assumed to be equivalent. The column for $r = 1$ in Table 4 has been worked out accordingly for $P_1(i) = .5$.

Next let us consider those exposures for which the first-order report is an unexposed syllable. By our assumptions, second-order reports will be determined by the same distribution of syllable strengths that determine first-order reports, save that the strength of the syllable emitted on first-order report must be zero. Suppose α to be the unexposed syllable emitted on first-order report. The second-order probabilities of all remaining syllables will differ from their first-order probabilities only by virtue of the fact that it is impossible for α to be emitted on second-order report, while the probability that α will be emitted on first-order report, $P_1(\alpha)$, is greater than zero. Consider a large number of exposures N for which there will be $n_i = NP_1(i)$ first-order reports of the exposed syllable and $n_\alpha = NP_1(\alpha)$ first-order reports of the unexposed syllable α. We now ask, what is the probability of i if all α responses are eliminated? The answer, $n_i/(N - n_\alpha)$, gives us the probability $P_2(i)$ that the exposed syllable will be given on second-order report if α is the first-order report. The second-order probabilities of each remaining unexposed syllable can be found in the same way. For the unexposed syllable β, for instance, $P_2(\beta) = n_\beta/(N - n_\alpha)$, where $n_\beta = NP_1(\beta)$. These probabilities of second-order report are illustrated in the column for $r = 2$ in Table 4. Since all unexposed syllables are equally probable, the same solution would be reached if the syllable emitted on first-order report were some unexposed syllable other than α. We can therefore let the general symbol j_1 represent whatever unexposed syllable is emitted on first-order report. It follows that

$$P_2(i) = \frac{n_i}{N - n_{j_1}} = \frac{NP_1(i)}{N - NP_1(j_1)} = \frac{P_1(i)}{1 - P_1(j_1)}. \quad [5]$$

The reasoning upon which this equation is based can be seen more clearly when stated in terms of a conventional statistical device. Suppose there to be an urn containing N balls which differ only with respect to color. There are balls of m different colors in the urn; n_i of them are white and the remaining balls are divided equally among the $m - 1$ remaining colors $\alpha, \beta, \gamma, \cdots$. After each draw from the urn we remove all balls having the color of the ball that was drawn. Suppose the initial draw turns up a ball of color α. All balls of that color are thereupon removed from the urn, leaving only $N - n_\alpha$ balls in the urn for the second draw. Then the probability of obtaining a white ball on the second draw will be $n_i/(N - n_\alpha)$. Since $n_\alpha = n_\beta = \cdots$, it can be said in general that the probability $P_2(i)$ of obtaining a white ball on second draw if the first draw yields a nonwhite ball is $n_i/(N - n_{j_1})$, where $n_{j_1} = n_\alpha = n_\beta = \cdots$ is the number of balls having the same color as the ball which was first drawn from the urn. Equation 5 follows immediately from the definitions $P_1(i) = n_i/N$ and $P_1(j_1) = n_{j_1}/N$.

The same reasoning can be followed to show that

$$P_3(i) = \frac{P_2(i)}{1 - P_2(j_2)}, \qquad [6]$$

where $P_2(j_2)$ is the second-order probability of whatever unexposed syllable is emitted on second-order report. Third-order probabilities calculated on the assumption that the first- and second-order reports are the unexposed syllables α and β, respectively, appear in the column for $r = 3$ in Table 4. In general, the probability $P_r(i)$ that the exposed syllable will be emitted on the rth-order report if unexposed syllables have been emitted on all preceding reports is

$$P_r(i) = \frac{P_{r-1}(i)}{1 - P_{r-1}(j_{r-1})}. \qquad [7]$$

For reports of a given order, say the rth, the probability of any unexposed syllable, $P_r(j_r)$ will equal the complement of the probability of the exposed syllable, $Q_r(i) = 1 - P_r(i)$, divided by the number of equally-likely unexposed syllables. On the rth order of report the number of unexposed syllables whose strengths are greater than zero will be $(m - r)$, since $(r - 1)$ of the original $(m - 1)$ unexposed syllables will have been reduced to zero strength by virtue of having been emitted on an earlier report. Then for the $(r - 1)$th report

$$P_{r-1}(j_{r-1}) = \frac{Q_{r-1}(i)}{m - (r - 1)} = \frac{1 - P_{r-1}(i)}{m - (r - 1)}. \qquad [8]$$

Substituting into Equation 7, we obtain

$$P_r(i) = \frac{P_{r-1}(i)}{1 - \dfrac{1 - P_{r-1}(i)}{m - (r - 1)}} = \frac{(m - r + 1)P_{r-1}(i)}{(m - r) + P_{r-1}(i)}. \qquad [9]$$

Similarly,

$$P_{r-1}(i) = \frac{P_{r-2}(i)}{1 - P_{r-2}(j_{r-2})} = \frac{(m - r + 2)P_{r-2}(i)}{(m - r + 1) + P_{r-2}(i)}. \qquad [10]$$

To obtain $P_r(i)$ in terms of $P_{r-2}(i)$ we substitute for $P_{r-1}(i)$ in [9] its value in [10] and simplify. This gives

$$P_r(i) = \frac{(m - r + 2)P_{r-2}(i)}{(m - r) + 2P_{r-2}(i)}. \qquad [11]$$

When this reduction is carried out a times,

$$P_r(i) = \frac{(m - r + a)P_{r-a}(i)}{(m - r) + aP_{r-a}(i)}. \qquad [12]$$

Proof of Equation 12 is by mathematical induction. The $(a + 1)$th reduction is defined by the solution of Equation 9 for $P_{r-a}(i)$ in terms of $P_{(r-a)-1}(i)$. Substituting for $P_{r-a}(i)$ in Equation 12 and simplifying, we have

$$P_r(i) = \frac{\dfrac{[m - r + a][m - (r - a) + 1]P_{(r-a)-1}(i)}{m - (r - a) + P_{(r-a)-1}(i)}}{(m - r) + \dfrac{a[m - (r - a) + 1]P_{(r-a)-1}(i)}{m - (r - a) + P_{(r-a)-1}(i)}}$$

$$= \frac{[m - r + a][m - r + a + 1]P_{r-a-1}(i)}{[m - r][(m - r + a) + P_{r-a-1}(i)] + a(m - r + a + 1)P_{r-a-1}(i)}. \qquad [13]$$

With further simplification the denominator becomes

$$(m - r)(m - r + a) + P_{r-a-1}(i)[(m - r) + a(m - r + a + 1)]$$
$$= (m - r)(m - r + a) + P_{r-a-1}(i)[(a + 1)(m - r + a)].$$

The $(m - r + a)$ terms cancel with the same term in the numerator of Equation 13, and, after bracketing,

$$P_r(i) = \frac{[m - r + (a + 1)]P_{r-(a+1)}(i)}{(m - r) + (a + 1)P_{r-(a+1)}(i)}. \quad [14]$$

The conditions for proof by mathematical induction are fulfilled, since Equation 14 is the equivalent of Equation 12 for $a = (a + 1)$.

The desired solution for $P_r(i)$ in terms of $P_1(i)$ is a special case of Equation 12. Let $a = (r - 1)$, then

$$P_r(i) = \frac{[m - r + (r - 1)]P_{r-(r-1)}(i)}{(m - r) + (r - 1)P_{r-(r-1)}(i)}$$
$$= \frac{(m - 1)P_1(i)}{(m - r) + (r - 1)P_1(i)}. \quad [15]$$

Equation 15 states the probability that the exposed syllable will be reported as at least the rth-order report *in terms of its probability as the first-order report*. It can therefore be of great value in the theoretical analysis of tachistoscopic experiments. It provides a direct and independent test of the assumptions from which the Lazarus and McCleary data have been deduced here.[2] What is of greater importance, it provides a quantitative, predictive test of experimental measurements of the stochastic variables that are assumed to underlie perceptual report. Such a formulation can be of value not only in the further analysis of subception phenomena, but in the development of a general theory of the role of response processes in perception.

It will be observed that Equation 15 depends upon the assumption that the various unexposed syllables have equal first-order probabilities. Actually the equation will give fairly accurate solutions when these probabilities are not equal, unless the differences are very extreme. Such cases are not likely to arise in experiments making use of nonsense syllables. If an exact solution should be desired under these conditions, it can be obtained by substituting experimental values for $P_1(j_1)$, $P_2(j_2)$, $\cdots$ $P_r(j_r)$ and carrying out successive reductions in the manner of Equations 9 and 10. The only additional data required by this type of solution are the first-order probabilities of all unexposed syllables.

Summary and Conclusion

Since the statistical assumptions put forward here account for the subception effect, it is unnecessary to postulate the existence of a subception process ("discrimination without awareness"). Although postulation of a subception process cannot be excluded as an alternative interpretation, the present assumptions can be shown to possess several advantages over it.

1. The present assumptions are consistent with what has been called the symbolic-report hypothesis. This hypothesis is a general principle of wide application to perceptual phenomena. As long as this hypothesis can be shown to hold without exception, it provides an empirical law of great analytic value.

2. From the present assumptions it is possible to deduce some 13 definite inequalities of GSR between the Lazarus and McCleary experi-

[2] Experimental proof that the probability of correct report is greater than chance appeared after this paper was completed (Bricker, P. D., & Chapanis, A. Do incorrectly perceived tachistoscopic stimuli convey some information? *Psychol. Rev.*, 1953, 60, 181–188). The data reported there, however, are not extensive enough for a quantitative test of Equation 15.

mental categories, whereas only one inequality follows rigorously from the subception hypothesis itself.

3. The present assumptions assign a definite significance to exposure time as a parameter of the inequalities of GSR. Exposure time will affect inequalities differently according to the categories being compared. Although the Lazarus and McCleary data are insufficient for this kind of analysis, the effects can be tested in further experiments.

4. The present assumptions yield, in addition to the interpretation of GSR inequalities, quantitative deductions of syllable probabilities for lower-order reports. The subception hypothesis offers no interpretation of these data.

5. The present assumptions are subject to quantitative formulation on the basis of experimental data. The two basic relationships needed are (*a*) the function relating average strength of a syllable process to the duration for which the corresponding syllable is exposed, and (*b*) the function relating mean GSR to the average strength of the processes underlying shock syllables. Since the average strength of a syllable process is defined in terms of the probability of that syllable's report, both of these functions can be determined by experiment.

All these advantages may be subsumed under the statement that the statistical form of the symbolic-report hypothesis possesses far greater potentiality for the analytic description of perceptual phenomena than does the subception hypothesis. Even if those implications which cannot yet be compared with experimental data should fail to be confirmed by later tests, this general advantage would still rest with assumptions of the type of the present ones. For only when *ad hoc* postulations like the subception hypothesis are replaced with theoretical concepts having the mathematical properties of variables can rigorous description be given to a wide range of phenomena.

REFERENCES

1. Howes, D. H., & Solomon, R. L. A note on McGinnies' "Emotionality and perceptual defense." *Psychol. Rev.*, 1950, **57**, 229–234.
2. Howes, D. H., & Solomon, R. L. Visual duration threshold as a function of word probability. *J. exp. Psychol.*, 1951, **41**, 401–410.
3. Lazarus, R. S., & McCleary, R. A. Autonomic discrimination without awareness: a study of subception. *Psychol. Rev.*, 1951, **58**, 113–122.
4. McGinnies, E. M. Emotionality and perceptual defense. *Psychol. Rev.*, 1949, **56**, 244–251.
5. McGinnies, E. M., Comer, P. B., & Lacey, O. L. Visual recognition thresholds as a function of word length and word frequency. *J. exp. Psychol.*, 1952, **44**, 65–69.
6. Preston, K. A. The speed of word perception and its relation to reading ability. *J. gen. Psychol.*, 1935, **13**, 199–203.
7. Solomon, R. L., & Postman, L. Frequency of usage as a determinant of recognition thresholds for words. *J. exp. Psychol.*, 1952, **43**, 195–201.

(Received April 1, 1953)

Psychological Review
Vol. 61, No. 2, 1954

REINFORCEMENT AND EXTINCTION PHENOMENA[1]

JACK L. MAATSCH

Michigan State College

This paper will be concerned with a redefinition of the concept of reinforcement and its application to a comprehensive interference theory of inhibitory phenomena. These two distinct positions are part of a general theory of behavior under development by the author. However, from the standpoint of the more conventional conceptions of learning, it is felt that these two positions may stand by themselves and serve as a partial theoretical background for a series of experimental studies (1, 11, 12, 14). These studies may be considered in the light of applications of the general theory and particularly of the positions here developed.

Abstraction and the Notion of Reinforcement

The notion of reinforcement is an abstraction. It is a construct which has a definite role in a system of abstract ideas (theory); and to be meaningful as an idea, or as a linguistic term denoting the idea entertained, the definiens must contain, or be reducible to, symbols denoting observables. If the idea is an abstraction from objects of the *thing* level, e.g., animal or wood, then the simple ostensive or denotative mode of definition would suffice. If, however, the abstraction is one pertaining to unobservables, as in the case of reinforcement, then the operational definition is the proper form of definition to secure intersubjective agreement concerning the possibility of entertaining the idea in question (13).

We may distinguish the *theoretical* properties of a construct (its functional relationship to other abstractions in a theory) from a given operational definition of that idea. The notion of reinforcement carries the connotation of cause.[2] Reinforcement causes learning. As a causal agent, then, the assertion or denial of the idea is tantamount to the prediction of learning or its absence in a given situation. Reinforcement, once asserted, must specify what is learned, that is, what response is learned to what stimuli. Therefore, we may expect that an adequate operational definition of the notion will determine the grounds for asserting or denying learning; and further, it will precisely specify what response is learned to what stimuli.

We wish to defend the theoretical idea of reinforcement; but we wish to deny contemporary operational definitions of

[1] I wish to express my gratitude for invaluable aid received through discussion with Professor M. Ray Denny, R. A. Behan, R. H. Davis, H. M. Adelman, O. Smith, R. Martindale, and M. Goer.

[2] The term "cause" is here used in the sense usually employed within philosophy of science. To assert that such and such a state of affairs *causes* learning, is to assert that there exists a more molecular chain of events (a neurophysiological chain of events) which links the observed "state of affairs" with the observed modification of behavior. Therefore, the definition of reinforcement should define the parameters necessary to produce a neurophysiological chain of events which makes reasonable the relationship between such and such a situation and the observed modification of behavior. Thus it may be argued that the notion of sheer "contiguity" specifies only one of the parameters for learning, that the notion of "drive reduction" specifies an erroneous, or at least unparsimonious, parameter, and finally, that "elicitation," the hypothesis to be proposed, does delineate an acceptable and parsimonious set of parameters.

this idea, specifically the "drive reduction" hypothesis and the "contiguity" hypothesis. The drive reduction hypothesis would seem to be empirically false (5, 6, 18, 19), neurologically implausible as the cause of learning (5, 6), and operationally inconsistent. For example, compare the definitions and applications of primary and secondary reinforcement (10). "Contiguity" as formulated by Tolman (22) and Guthrie (4) is universally applicable and therefore meaningless within the present system. One may never *deny* the occurrence of learning to contiguous stimuli; therefore strict adherence to a contiguity hypothesis yields false predictions when utilized in conjunction with the habit concept. One must predict that the animal will tend to learn to do what it did on the previous trial. Thus, a redefinition of the construct reinforcement would seem to be in order at the present time.

THE ELICITATION HYPOTHESIS

A reinforcement (E) will occur whenever there occurs a stimulus (s) or a stimulus complex (S) that elicits a characteristic response (r). Given the occurrence of a reinforcement, there will result an increment to a tendency ($_SH_R$) for that complex to evoke a member of that response class (R).

There are several features of the definition which require some elaboration. First, the critical operation to secure or to assert a reinforcement is that of producing or causing some identifiable member of a response *class* to occur. For example, the experimenter makes an animal hungry so that it will approach and eat a pellet of food. Secondly, a member of *that* response class will reoccur with increasing probability and decreasing latency, and a member of that response class will occur if stimulus generalization occurs in a behavior sequence. The anticipatory occurrences of responses which are produced by stimulus generalization of the habit to places temporally antecedent to the eliciting stimulus complex may provide the basis for the explanation of extinction phenomena. This matter will be discussed below.

The operational determination of the reinforcing stimulus, where seemingly indeterminant, e.g., the latent learning catastrophe, may proceed by the time-honored methods of determining causal agents if one is wary of the inferential pitfalls of such methods (3, ch. 13). Within the present formulation of reinforcement, to test for latent learning in an experimental situation that excludes reinforcement, one merely eliminates the stimulus or stimulus complex that produces the response class to be "latently" learned.

One additional point concerning the definition should be made, namely, the lack of an operational distinction between primary and secondary reinforcers. Both elicit identifiable responses; hence both reinforce. One type is prior to and not dependent upon previous learning, e.g., shock, while the other *is* contingent upon previous learning, e.g., discriminated stimuli. There is, however, a difference in the effects of using these different types of reinforcing agents to produce new learning. This leads us to a third type of reinforcer which is also contingent upon past learning and upon the presentation or the occurrence of "secondary" reinforcers. We will tentatively define frustration as follows: *Frustration stimulation* (s_f) will occur in a learned sequence whenever the elicitation of a learned response (CR) results in the occurrence of a stimulation complex (S) interrupting performance of the learned sequence. s_f will elicit members of a characteristic class of responses (Rs_f).

The definitions of E and s_f, although structurally similar, are of a quite different nature with respect to the abstract idea entertained. Frustration is a particular kind of stimulation occurring in a specific type of situation; and learning with respect to the situation must have occurred to some extent before s_f will occur. Care should be taken to distinguish s_f from emotion, here considered an unconditioned response to other types of situations, although frustration may result in "emotional" behavior. Further, s_f has the property to *elicit* a class of characteristic behavior. Therefore, the construct may be considered an axillary of E, specifically *demanding* the assertion of E. Since frustrating situations are reinforcing, the stimulus complex which initiates s_f will be conditioned to Rs_f.

The reactions to s_f will vary as the particular physical situation varies. For example, s_f occurring after an anticipatory turn into a cul results in responses characterized by "recoil," whereas s_f occurring in an enclosed goal box results in behavior grossly described as "emotional." Habits based upon s_f behave as any habit and interact with the original habit frustrated. It is this interaction of habits based upon two different sources of reinforcement, e.g., *US*, s_f, which seems to suggest an explanation of some complex extinction phenomena.

Applications

Since the selective marshaling of positive experimental evidence is not particularly impressive or parsimonious, perhaps it will be better to apply the hypothesis in a few general cases.

Shock and learning. In experiments involving shock, extreme care should be taken to observe exactly what kind of responses are actually being elicited from each animal. These are the responses that will be learned. For example, consider the experiment by Brogden, Lipman, and Culler (2) which has played an important role in the development of the two types of conditioning formulated by Hilgard and Marquis (7). Under the classical procedure, animals are shocked with every presentation of a buzzer while in a revolving cage. The results of this procedure show, first, a steady but slow increase in "anticipatory running," and then, in the latter stages of the experiment, a steady decay in anticipatory running. On the other hand, during instrumental escape administration, shock is withheld if the running response appears in anticipation of administration of shock. Here the results show rapid, unhindered learning.

Consider the effect of the two methods of administration of shock upon the responses elicited (17). With the classical procedure, the running response is elicited during the early stages of the experiment, for the animal is relatively stationary. As this response begins to appear in anticipation of shock, the shock appears while the animal is in motion. The response elicited, if anything, disrupts or interferes with the previously learned running response. Thus, the classical procedure, which does not assure consistent effects of shock throughout the course of learning, results in first an initial rise in the learning of "anticipatory running" and then gradual disruption of the response to be learned. Of course, new responses are now being elicited in anticipation of shock. On the other hand, with the instrumental procedure, shock is withheld if the running response is elicited by the buzzer. Thus, the response is protected from disruption by shock. Unhindered learning does result with the instrumental procedure. From the present position, the distinction between the two kinds of learning with respect to shock would seem to be theoretically superficial.

Two-factor theories of reinforcement. The conditioning of internal and autonomically produced reflexes must be treated somewhat differently from the conditioning of skeletal responses. A characteristic of internal reflexes, not shared by the instrumental skeletal response, is that of being *elicited* by the initial phases or *onset* of the unconditioned stimulus and of being maintained throughout the duration of shock. If shock persists for a period of time, the identical internal reflex persists, while a wide variety of skeletal responses may appear throughout the duration of shock. Now if the external stimulation complex is markedly changed *during* the continued presence of shock, the stimulus complex associated with onset of shock may acquire the property to elicit internal responses but not the ensuing instrumental escape response. These internal reflexes would, however, form a distinct and persistent internal stimulation pattern which would be part of the total stimulation complex eliciting the skeletal response instrumentally made to escape shock. Therefore, with the continued repetition of the sequence, elicitation of internal responses without the elicitation of the instrumental skeletal response would become increasingly improbable.

When skeletal responses are made to continued presence of shock a different hypothesis emerges. A response elicited by a given stimulation complex is a response reinforced to that complex. However, a second and different response elicited to the same external complex would also be reinforced, the second response replacing the first as the response most likely to occur subsequently if the shock and thus the sequence of events were terminated. In other words, the last response elicited by a given stimulation complex will be the first to occur to that complex on subsequent occasions. This principle was first applied by Guthrie (4) and will be used in a similar manner in applying the elicitation hypothesis.

Through the above analysis it is seen that "drive reduction" as presently applied within the Hullian framework has only a tenuous correlation with the response to be learned. When considering the differences in the analysis of the learning of the two types of responses, internal and skeletal, it becomes apparent that two types of learning (7) or two types of reinforcement (15, 20) could easily be posited. "Contiguity" with the onset of noxious stimuli or "contiguity" with the elicitation of *internal* responses becomes a special case of the application of the elicitation hypothesis as does the reinforcement of the terminal *skeletal* response through "drive reduction."

Classical conditioning. The learning of a behavior sequence originates with the involvement of an *S* complex with a reinforcing stimulus. The first influence of a reinforcement is upon stimuli coexisting spatially and temporally. The effect of a reinforcement is upon a habit involving that complex with the response class produced. Thus, irrelevant (conditioned) stimuli come to stand in the same relation to response as does the unconditioned stimulus. This is of course similar to the substitution hypothesis as elaborated by Hilgard and Marquis (7). Because of reinforcement of trace elements present in the eliciting complex and the similarity of internal and external stimulation, i.e., hunger pangs, gray alley and goal box, constant illumination, etc., on subsequent trials, spatially and temporally prior complexes will elicit R_o. In classical conditioning, this response class is the molecular internal unlearned salivary reflex elicited by meat powder, while the temporally antecedent complex involves a ticking metronome as a discriminative stimulus.

Natural approach and avoidance. The analysis thus far is also descriptive of the learning of natural sequences which terminate with pleasant or noxious unconditioned stimuli. On subsequent trials, S's antecedent to S_o (the eliciting complex) come to elicit R_o (the elicited response class). In case R_o is an approach class, S's antecedent in time to S_o elicit an approach response which results in the next S complex which will in turn elicit approach responses. This sequence of events if not interrupted will automatically result in S_o and termination of the sequence. If R_o is an avoidant class, then temporally antecedent S's will elicit responses which do not lead to the complex more approximate to S_o. This sequence of events, if unhindered, automatically leads to the elimination of the sequence.

An interference theory of inhibition. This position postulates that extinction is a result of the learning of responses which are elicited by the omission of the reinforcing stimulus in a learned situation. This new response tendency interferes with the original response tendency. Complete extinction, then, is viewed as an unstable equilibrium of two competing response tendencies to a common stimulus complex (11).

By applying Hull's stimulus-patterning corollary V (9, p. 379) to the extinction situation, the following law may be derived: any change in the stimulus situation which was present during learning will result in the learning of competing responses (Rs_f) to a discriminative stimulus complex (Ss_f). Where minimal change of the stimulus complex occurs, e.g., omission of reward only, extinction will be described as *relearning*. Where marked changes occur at the beginning of extinction, we will refer to the extinction process as *discrimination learning*. It follows from corollary V (9, p. 368) that the greater the stimulus difference between the learning and extinction situations, the faster the discrimination and, as a consequence, the faster the so-called "inhibition" of the original response due to learning of an incompatible response to frustration. The introjection of massing procedures when learning was spaced or the elimination of "secondary reinforcing" (discriminative) stimuli may serve as an example of this prediction.

In the following section, we should like to try to answer specifically the difficulties facing the formulation of this type of theory as posed in a review of theories of learning:

> 1. . . . what makes the competing response occur initially, and how does it eventually become stronger than the originally learned response? (21, p. 713.)

The competing response is initially elicited by frustration resulting from the removal of the reward in the learned sequence. This state of affairs, by definition, results in the reinforcement of "that response" to "that stimulus complex." During extinction, as long as the original response tendency is stronger, frustration will again occur and the competing response tendency will be further strengthened. An unstable equilibrium is the end result of "complete" extinction with the new interfering tendency slightly the stronger of the two tendencies.

> 2. One of the most embarrassing findings is the fact that the massing of trials has opposite effects on conditioning and extinction (21, p. 713).

Original conditioning usually concerns the development of habits to new stimuli in an unfamiliar situation, while extinction involves the learning of new responses in a previously learned situation. As a consequence, care should be taken when generalizing from the effects of variables in the one situation to effects of the variable in the other situation. Massing procedures often generate unconditioned emotional responses. These would tend to interfere with the learning of a new response. However, during extinction these same emotional responses would facilitate the extinction of the learned response. The effect of massing procedures upon rate of extinction should also depend upon the kind of procedure employed during learning. When learning is spaced, and it usually is, massing of extinction trials results in a discriminative extinction situation, and extinction of the original response should be facilitated.

3. . . . the absence of positive correlations between conditioning and extinction measures . . . (21, p. 713).

In order to derive positive correlations between the rate of learning and the rate of extinction utilizing an interference theory it is necessary to assume: (*a*) that the two situations are identical *and they are not here considered to be*, (*b*) that there exists some general "learning ability," which does not seem to exist at all (8, p. 635), and (*c*) that this ability is the *only* variable contributing to rate of learning and extinction, which is at least questionable. However, we might assume that rate of learning is indicative of adjustment to the experimental apparatus and procedure which is in turn related to some emotionality factor. Here we have a general factor common to both learning and extinction. However, this factor could result in slow learning but fast extinction since the occurrence of emotional responses would result in additional interference to the original response. In this case the interference theory would generate the prediction of negative correlations. In general, the absence of positive correlations may be viewed as a confirmation of, and not an inconsistency with, predictions stemming from a pure interference theory.

4. . . . the differential effects of an extra stimulus on responses extinguished to different degrees . . . (21, p. 713).

The phenomenon of disinhibition seems to be a rather complex phenomenon and offers difficulty to rival theories of inhibition (20, pp. 96–102; 21, p. 715). However, within the present position, if we can assume that the extra stimulus is sufficiently pervasive to disrupt the stimulus complex actually eliciting the response class in question, several predictions can be made. First, in a relearning type of extinction situation, since the complex in question elicits both the learned and interfering habits indiscriminately, little or no disinhibition would be manifested. In the discriminative type of extinction situation, the extra stimulus would selectively disrupt the discriminative stimuli to which the interfering responses are being learned, allowing the original habit to manifest itself more rapidly following presentation of the extra stimulus. Granted the above analysis, we may expect in the relearning type of situation that any observed disinhibition would decrease as extinction progressed. In the discrimination type of situation, we might expect the opposite. Suffice it to say that a pure interference theory may be able to handle adequately the phenomenon of disinhibition; however, much more research in this area is needed before the phenomenon of disinhibition may be regarded as embarrassing to any theory of inhibition.

5. . . . the different types of extinction curves following massed and distributed conditioning . . . (21, p. 714).

In general, the original learning conditions and their relationship to the conditions obtaining during extinction are considered the primary determiners of the obtained extinction curves, and variations in the learning conditions will certainly produce noticeable effects upon extinction curves, e.g., partial reinforcement, variation of intertrial interval, etc.

6. . . . the extinction of a conditioned pupillary response in a curarized cat (21, p. 714).

It would seem plausible to attribute extinction of this reflex to physiological fatigue since the response in question involves such a limited number of antagonistically operating muscles. The interference theory here employed is concerned with the prediction of gross skeletal responses of intact organisms in complex stimulus fields under normal learning conditions. In this case, physiological fatigue or inhibitory states arising from performance or work are considered irrelevant variables. This is not to deny that physiological fatigue is irrelevant, but to assert that its effect in the usual learning experiment is negligible. For experimental evidence for this assumption see (12). A more interesting problem might be to explain the learning of this response from the drive reduction position of reinforcement.

7. . . . and even the phenomena of spontaneous recovery . . . (21, p. 714).

Spontaneous recovery within this system is considered to be a function of the reinstatement of cues previously associated with reward. Since extinction is described as an unstable equilibrium between competing response tendencies, any partial reinstatement of cues which have been discriminatively associated with reward may be expected to result in a disequilibrium and consequently "spontaneous" recovery. The theory affords a system in which many predictions concerning spontaneous recovery may be generated, some of which contradict predictions stemming from the application of the theory proposed by Hull.

8. Still other difficulties confronting this type of interference theory may be found in Razran's thorough discussion of the problem (21, p. 714).

An analysis of Razran's (16) criticisms, which are experimentally founded and are not considered above, yielded no serious difficulties for the present position within the scope intended.

Having specifically answered a list of traditional criticisms of pure interference theories, I would like to summarize briefly a few of the more important experimental results we have obtained directly as a consequence of testing predictions stemming from the applications of the theoretical position elaborated above.

1. The massing procedure serves as a cue when discriminately associated with nonreward (11).

2. Response inhibition may be rapidly learned where no previous inhibition (I_R) was present (11).

3. Amount of effort expended during extinction of a molar response is not a significant factor in the extinction process (12).

4. Partial reinforcement produces discrimination learning, and the resulting resistance to extinction is a function of the degree of discrimination obtained prior to extinction, *OTE* (14). The tentative generic formula suggested by the theory for fixed ratio reinforcement in a Skinner-box situation is as follows: The number of responses to extinction resulting from fixed-ratio reinforcement equals the percentage of discrimination obtained under partial reinforcement prior to extinction times the reciprocal of the ratio of reinforcement employed times the number of responses to extinction of a maximal habit reinforced at a ratio of 1.0 (100%). Percentage of discrimination is defined as the ratio of responses to the food dish to the number of bar presses per reward, $D = 1 - \frac{A-1}{N-1}$.

5. Partial reinforcement per se does not yield increasing resistance to extinction where discrimination is impossible (14).

6. Resistance to extinction of a running response is a function of the type of response elicited by frustration (removal of reward). A directly opposing or incompatible response to frustration yields rapid extinction of the original response, ***while a response compatible***

with the original running response yields little or no extinction of the original response (1).

7. The response to frustration will not extinguish without the introjection of some other frustrating stimulus blocking completion of that response (1).

Statements 6 and 7 when considered together yield a theoretical prediction and an experimental confirmation of the relation between fixation and extinction of a response tendency. A series of predictions drawn from the general theory concerning fixation are presently being investigated, with the hope of very simply demonstrating a "functionally autonomous" maladaptive running response in the rat.

SUMMARY

This paper has attempted to formulate a new reinforcement hypothesis and to defend a pure interference theory of extinction. We have attempted to indicate the parsimonious integrative power of the elicitation hypothesis and its role in producing inhibition in the hope that specific and detailed experimental evidence soon to be presented will be more meaningful since the facts obtained represent somewhat of a departure from hypotheses drawn from conventional behavior theory.

REFERENCES

1. ADELMAN, H. M., & MAATSCH, J. L. Resistance to extinction as a function of the type of response elicited by frustration. *J. exp. Psychol.*, in press.
2. BROGDEN, W. J., LIPMAN, E. A., & CULLER, E. The role of incentive in conditioning and extinction. *Amer. J. Psychol.*, 1938, 51, 109–117.
3. COHEN, M. R., & NAGEL, E. *An introduction to logic and the scientific method.* New York: Harcourt, Brace, 1934.
4. GUTHRIE, E. R. *The psychology of learning.* New York: Harper, 1935.
5. HARLOW, H. F. Mice, monkeys, men, and motives. *Psychol. Rev.*, 1953, 60, 23–32.
6. HEBB, D. O. *The organization of behavior.* New York: Wiley, 1949.
7. HILGARD, E. R., & MARQUIS, D. G. *Conditioning and learning.* New York: D. Appleton-Century, 1940.
8. HOVLAND, C. I. Human learning and retention. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 613–689.
9. HULL, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
10. HULL, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.
11. MAATSCH, J. L. An exploratory study of the possible differential inhibitory effects of frustration and work inhibition. Unpublished master's thesis, Michigan State College, 1951.
12. MAATSCH, J. L., ADELMAN, H. M., & DENNY, M. R. Effort and resistance to extinction of the bar-pressing response. *J. comp. physiol. Psychol.*, 1954, 47, 47–50.
13. MAATSCH, J. L., & BEHAN, R. A. Toward a more rigorous theoretical language. *Psychol. Rev.*, 1953, 60, 189–196.
14. MAATSCH, J. L., DENNY, M. R., & WELLS, R. H. Resistance to extinction as a function of the number of blocks of fixed ratio reinforcement. *J. exp. Psychol.*, in press.
15. MOWRER, O. H. On the dual nature of learning—a reinterpretation of "conditioning" and "problem solving." *Harvard educ. Rev.*, 1947, 17, 102–148.
16. RAZRAN, G. S. The nature of the extinction process. *Psychol. Rev.*, 1939, 46, 264–297.
17. SHEFFIELD, F. D. Avoidance training and the contiguity principle. *J. comp. physiol. Psychol.*, 1948, 41, 165–177.
18. SHEFFIELD, F. D., & ROBY, T. B. Reward value of a nonnutritive sweet taste. *J. comp. physiol. Psychol.*, 1950, 43, 471–481.
19. SHEFFIELD, F. D., WULFF, J. J., & BACKER, R. Reward value of copulation without sex drive reduction. *J. comp. physiol. Psychol.*, 1951, 44, 3–8.
20. SKINNER, B. F. *The behavior of organisms.* New York: Appleton-Century-Crofts, 1938.
21. SPENCE, K. W. Theoretical interpretations of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 690–729.
22. TOLMAN, E. C. *Purposive behavior in animals and men.* New York: Appleton-Century-Crofts, 1932.

(Received May 19, 1953)

Psychological Review
Vol. 61, No. 2, 1954

HYPOTHETICAL CONSTRUCTS AND INTERVENING VARIABLES [1]

ARTHUR GINSBERG

New York University

As MacCorquodale and Meehl correctly maintain, "the view which theoretical psychologists take towards intervening variables and hypothetical constructs will of course profoundly influence the direction of theoretical thought" (14, p. 95). Accordingly, it is important to examine whether the grounds for this distinction, as formulated by these authors, can be reasonably sustained, especially in face of mounting criticisms from various quarters (2, 12, 15, 16). The conclusions which the present paper will endeavor to justify are (*a*) that there are fundamental methodological and logical distinctions between what might be termed intervening variables and hypothetical constructs, but (*b*) that the grounds for these distinctions are other than those cited by MacCorquodale and Meehl. Some of the arguments presented by these authors seem to suffer from lack of clarity, and it may be because of this that the distinction has been repudiated and, even where accepted, not always with signs of adequate understanding.

Some Distinctions between Laws and Theories

1. Elsewhere (6), in somewhat closer detail, some of the characteristics distinguishing scientific laws and theories were outlined and supported against views which either denied the distinction or its importance or which maintained that laws only were the proper form which psychological knowledge should assume.

It is generally accepted that the central objective of scientific endeavor is to arrive at an understanding of the universe or various of its parts and facets. What does it mean to understand? This question is directed not upon things understood, but once removed, upon understanding itself. Its answer would afford an understanding of understanding. What are the methods, objects, and objectives of such once-removed understanding? Is it the behavioral processes undergone or methods employed by those questing understanding which is the object? Or is it the final product of such questing, the linguistic systems or logical artifacts arrived at? In either case, which behaviors or methods or which artifact outcomes are to be taken as paradigms? Connected with the foregoing, are the second-order methods involved in examining first-order methods of understanding or their outcomes identical methods, and are their outcomes of identical status or kind? That is, are the former part and parcel of a more inclusive and unitary enterprise of understanding or something quite apart and presupposed or immune? Are these second-order methods purely analytical, or are they empirical, or, perhaps, some combination? And if a combination, in what proportions and mode? What objectives are to be realized by such inquiries into the forms or tactics of knowing? Are these limited to description or may they be reconstructive in performing critical or legislative functions? And if the latter, by what warrant and to what degree?

[1] The writer is grateful to Profs. Ernest Nagel and Rita W. Nealon for critically reviewing the manuscript and offering suggestions toward its improvement.

These sundry questions suffice to illustrate the inevitability and (perhaps) perpetuity of divergent opinions concerning the meanings to be attached to and the answers to be settled upon the question "what does it mean to understand?" But it is an issue which must be joined and one which can result in enlightenment. Opinions are not all of one quality. They can be sorted out and appraised in respect to (*a*) their pragmatic significance in expediting the primary scientific objective, the understanding of things, (*b*) the degree to which they accord with the history of scientific development and the totality of common human experience, and (*c*) their measure of internal consistency and comprehensiveness.

To understand, as here to be understood, means to be able to explain or to be capable of rendering anything intelligible. Essential for any explanation is the utilization of a principle or set of principles as a major premise in an argument containing a statement or statements of fact and from which a conclusion is derived. The conclusion is explained if it can be demonstrated to be a logical consequence of the argument and if the premises are empirically certified to be true. There are two essential kinds of principles, those which are called laws and those which are called theories. The differences between them are crucial to an understanding of what it means to understand and to the distinction between intervening variables and hypothetical constructs to be advanced.

2. A scientific law is a universal, synthetic proposition whose terms, denoting abstract classes, are connected in some form of invariant association. Being synthetic means that a law refers to the universe and is thus susceptible of empirical confutation. Being universal means that a law applies without restriction to time or place: if it is true, it is true always and everywhere that that to which it pertains may exist. Accordingly, laws are extremely vulnerable, for their survival depends upon the absence of a single contrary instance. However, this vulnerability is only a surface weakness since the survival of a law through manifold tests confers a high degree of certitude upon it. Were laws in principle difficult to disprove, they could not warrant much confidence, and any law in fact so obtuse or resistant, is deservedly regarded with skepticism.

The concepts associated within a law are experimental in the sense of being explicable in terms of or directly bound up with observation sentences or protocol reports. Certification or confutation is relatively direct and unequivocal, if not in actual practice, then conceptually. But this directness and affiliation with observation is not solely a logical matter. It pertains to genetic considerations as well, since laws are commonly induced, whether by curve fitting or by a more liberal type of examination of accumulated data.

By way of illustration, consider Boyle's law, usually stated as $PV = RT$ (where R is a constant whose value depends upon the mass of the gas). Pressure, volume, and temperature are concepts which can be rigorously determined by relatively direct observational and mensurational techniques, and what they refer to is capable of being perceived. The law was adduced by Boyle as a result of manipulations upon specimens of gas whereby he noticed that with the contraction of the volume there occurred a correlated increase in the temperature and pressure in fixed ratios expressed by the law. It is a universal proposition because it applies to anything, anywhere, and at any time which happens to be a gas. That something is a gas can be readily ascertained by invoking other criteria involving laws, e.g., the effects of passing an electrical current through it, its spectrum, its viscosity, etc. That Boyle's law is synthetic is attested to by its susceptibility to confutation; that it is true, by the absence of such confutation despite repeated exposure to a variety of tests, by its consonance with other principles, and by its practical utility.

The law signifies that alterations in at least one of the properties comprehended must result in correlate alterations in the others in predictable proportions. The actually observed instances of this invariant association are explained by showing them to comprise a subset of a more inclusive range of permissible values. And we understand the mechanical behavior of a concrete instance of gas when we are capable of providing such an explanation.

3. In contrast to a law, a theory is not a product of inductive generalization nor is it constituted by concepts which are explicable in observation terms alone. Theories are *invented*, highly abstract and general systems of concepts whose association with observations or facts is indirect rather than explicit. No amount of data inevitably determines a theory in the same way as it determines a law. Presuming the requirement of maximal mathematical simplicity, a finite set of data can be so extrapolated and interpolated as to limit the possible form or expression of the law to a few or even one case. For this reason, laws are generally highly resistant to change, a fact which is amply borne out by the history of science. It is quite otherwise with theories. Because their association with observations or facts is relatively remote, theories are notoriously mutable in time (in greater or lesser degree), and at any one time, there is likely to exist more than one theory for a given domain of subject matter. For example, in order to rectify certain deficiencies in Planck and Bohr's original quantum theory, Heisenberg, de Broglie, Schrödinger, Dirac, and recently Einstein have formulated alternate theories all of which are mathematically equivalent but differ conceptually. Such a state of affairs should be reassuring to the psychologist who is disturbed because more than one theory of learning is presently contending for general approval.

Typical theories like those of Newtonian mechanics, Einsteinian relativity, Planck's quantum, or Maxwellian electromagnetics contain ideas which are speculative in the sense of involving a constructive imaginative grasp of some root analogy, universal attribute, or system of entities of suprasensible purport or character. Such concepts apply in a variety of situations which, before the theory was invented, appeared to be independent of one another. Because of their abstractness, these concepts are differently specialized in different contexts of inquiry concerned with different phenomena. For example, in mechanics, the concept of force functions in such varied contexts as Hooke's law, Archimedes' principle, the motion or attraction of bodies in space, the law of the pendulum, laws of vibration, etc., all involving different coefficients and requiring different specializations of "force." But however manifestly different, these phenomena are revealed as instances of a common pervasive natural order by virtue of their being explained in terms of a common theory.

Distinctively theoretical concepts are postulated jointly as a system where rules of combination and transformation allow for the logical derivation of hypotheses capable of empirical test. When the boundary conditions are appropriately specified, Boyle's law can be thus derived from the kinetic theory of gases. The concepts of this theory, e.g., aggregations of molecules behaving in mechanical fashion, are not themselves capable of being defined in terms of observations or translated into a sense-data language. Rather, it is via such derivations as Boyle's law that the theory is empirically tested or coordinated. Owing to its logical structure, a theory may be envisaged as a matrix of concepts within which myriad channels of progression, whether lateral or hierarchical, are provided which connect diverse facts or relations within a unitary schema. Their tremendous power for organizing experiences in an intelligible and elegant manner makes theories the supreme objective of scientific inquiry, the most penetrating and satisfying form of understanding.

4. The proper grounds for distinguishing intervening variables and hypothetical constructs derive from the differences between a law and a theory. In what follows, the fundamental idea to be advanced is that intervening variables refer to concepts which are defined in terms of or comprise the ingredients of laws, whereas hypothetical constructs refer to concepts which are the constituents of a theory. The former, therefore, are capable of ostensive or extensive formulation while the latter are either primitives within a theory and thus defined intensionally, i.e., in connection with the rules of the theory, or high-level derivatives of the primitives. Considered in this light, many of MacCorquodale and Meehl's characterizations receive a more coherent and cogent grounding. But it also follows that some of their arguments cannot be sanctioned.

The criticisms of MacCorquodale and Meehl's views that will be developed will function as a means for amplifying these contentions. Because of the obscurities present in their article, MacCorquodale and Meehl are open to a number of alternative interpretations, some of which are untenable.

Do Intervening Variables Denote?

There are a number of arguments in MacCorquodale and Meehl's article (14) which indicate that as they conceive the distinction, intervening variables do not denote whereas hypothetical constructs do. The following quotations prove this supposition: hypothetical constructs "involve the hypothesization of an *entity, process,* or *event*" whereas intervening variables are simply "operational" (pp. 95–96); intervening variables "are merely names attached to certain convenient groupings of terms in [an] empirically fitted equation . . . [and] 'exist' [only] in the trivial sense that the law holds" (p. 99); our notion of intervening variables "involves nothing which is not in the empirical laws that support them" (p. 100); intervening variables are strictly mathematical except "where the verbal accompaniment of a concept . . . makes it a hypothetical construct" (p. 101); an intervening variable is "simply a quantity obtained by a specified manipulation of the values of empirical variables" (p. 103); in contrast to an intervening variable, "it is the business of a hypothetical construct to be 'true'" (p. 104); an intervening variable "is merely a shorthand summarization [while] for hypothetical constructs, there is a surplus meaning that is existential" (p. 106).

Is this a defensible way of distinguishing between classes of concepts employed in science? Are there legitimate and useful terms in science which refer to nothing or which denote empty classes?

Clearly, by definition, logical terms are nonfactual in this sense and so far as they are found in laws or theories, they fulfill a purely syntactical function. Among these are such terms as are determined by or name the choice of measuring units; formation signs such as universal or existential operators, e.g., all, some, every, etc.; connectives, e.g., and, or, implies, etc.; and punctuation marks. All these signs are linguistic conventions applied in accordance with logical-syntactical employment rules and refer to nothing which is or might be existential. It is clear that intervening variables are not intended by MacCorquodale and Meehl to represent such nondenotative terms.

Such terms as "unicorn" or descriptions like "the present king of France" lack denotative meaning since there is no class of actual existences to which they would correctly apply. However, it is not inconceivable that there should have been or might yet be a time when these terms would correctly apply to a class of existences. At present, while assertions involving these terms are either false or without practical scientific consequences, they are not logical contradictions. Accordingly, we must allow that were certain conditions to obtain, e.g., if at time *t* and place *p* one could observe an animal shaped like a horse, with a long horn in its head, fleet of foot, etc., such terms might correctly apply. We deny them denotative meaning because such meaning attribution would flout attested facts or certified laws or theories.

It is clear that MacCorquodale and Meehl do not suppose intervening variables to be analogous to or instances of fictive terms. For example, ${}_SH_R$ is interpreted as an intervening variable. Yet, were it fictive in the sense of denoting nothing or in naming an entity, process, or property whose existence could not be truthfully affirmed in any case, we would have to allow that, so far as we understood the conditions whereby ${}_SH_R$ *would* apply correctly, were these conditions actually realized, the term would acquire denotative meaning. In a certain sense, this possibility is provided for. MacCorquodale and Meehl contend that when such terms as ${}_SH_R$ are provided with a physiological interpretation, they are converted into hypothetical constructs and are therefore either true or false, i.e., they acquire denotative meaning.

It does not seem likely that MacCorquodale and Meehl would maintain that no psychological concepts denote or that, unless they were somehow connected with physiological references, psychological terms would refer to nothing existential. Yet, at least on the issue of denotative meaning, it is difficult to see why ${}_SH_R$ should be different from terms like "perception" or "memory" and the like.

The suggestion that the distinction between intervening variables and hypothetical constructs be based on the latter's representing or connoting neural processes while the former do not, has been made by Hilgard (8, p. 265) and by Tolman, at least in his latest writings (17). As thus construed, the distinction is not only a logical one, but is also based on factual considerations. From the standpoint of the distinction as adumbrated here, that the differences between intervening variables and hypothetical constructs stem from the differences between laws and theories, whether or not an intervening variable or a hypothetical construct has physiological bearings is entirely irrelevant. There can be physiological intervening variables and physiological hypothetical constructs just as well as psychological ones so far as there exist physiological laws and theories and psychological ones.

However, and more to the point at issue, whether any or all psychological concepts could be reduced to or somehow coordinated with a set of physiological concepts or with a physiological theory would not signal the acquisition of denotative meaning by terms which previously lacked it. The reverse would be closer to the truth, i.e., that a term can be interpreted within an independently developed system would reinforce our warrant in believing the term to be factually significant.

The difference between (*a*) psychological concepts which might be or have been reduced to or coordinated with a set of physiological concepts or absorbed within a more inclusive unitary discipline and (*b*) psychological concepts which have not been or may be resistant to such convergence is not that the former have and the latter do not have a surplus meaning which is existential. To begin with, the possibility of such convergence is not a function of psychological concepts alone but also of the state of physiological knowledge or of scientific knowledge in general. Next, even where the connotation of a psychological concept is not sufficiently developed to possess physiological purport, it does not follow that the concept lacks any reference whatsoever. Thus, ${}_{S}H_{R}$ may be supposed to lack physiological purport and still denote a dispositional property or state of organisms.

This latter confusion appears to rest upon the nominalistic notion that only concrete properties or particulars are real or ostensively significant. However, abstract terms are not nonreferential. Such terms as "blackness," "docility," etc., and all concrete terms capable of having a suffix like "ness" or "ity" attached to them or of functioning in similar ways in discourse, name what some other term connotes. For example, as Lewis points out, " 'roundness' names that character or property which is essential in order that the term 'round thing' should apply. . . . For every concrete term, 'C,' there is a cognate term which denotes the significate of 'C' " (the essential property in virtue of which C is 'C') (13, pp. 41–42). These considerations also apply to abstractions which denote kinds of relations, such as belongingness or affinity, and to those which represent interpreted mathematical equations, and to dispositional properties or states.

There is a final class of terms which is often considered to be analytic, namely, terms comprising a proposition which, in a given inquiry context, functions as a nominal definition or terms which function as synonyms for other terms. Since such terms assert nothing, they contribute nothing new or significant to scientific discourse save toward achieving clarity and economy of expression. According to Hempel, "a nominal definition singles out a certain *concept,* i.e., a nonlinguistic entity such as a property, a class, a relation, a function, or the like, and, for convenient reference, lays down a special name for it" (7, p. 4).

Nominal definitions are always conventional since they are rules for the use of language involving a stipulation that a *definiendum* be synonymous with a *definiens.* However, once the *definiendum* is so defined that instances of it can be readily identified, it may function in other contexts as an explanatory or factual concept. Thus if we suppose the mathematical statement of Hull's postulate 4 (10, p. 178) to be a nominal definition such that ${}_{S}H_{R}$ functions as a synonym for $M\ (l - e^{-kw})\ e^{-jt}\ e^{-ut'}\ (l - e^{-iN})$, it is not the case that ${}_{S}H_{R}$ also functions analytically in the other postulates of the system. On the contrary, all other appearances of ${}_{S}H_{R}$ are factual, and so when it occurs as a variable in the formula for excitatory potential (${}_{S}E_{R}$), it must possess denotative meaning. In a word, which formula in the system is taken as defining the term may be arbitrary, but once so defined, the other instances are strictly factual. Except for the classes of logical terms noted, no other significant term in an empirical theory or system of laws *always* functions analytically or without denotative meaning; i.e, no term for which independent criteria of application exist is analytic in every context whatsoever.

A term may be said to have meaning in a number of senses or modes. (*a*) A term *denotes ostensively* when it refers to a determinate particular and *extensively* when it refers to a determinate class where assertions about the particular or class are true or where these exist. (*b*) A term *connotes* in four ways: (*i*) the *conventional* connotation of a term consists in stipulations concerning its proper usage in discourse, i.e., that meaning which is generally accepted as fixing reciprocity of understanding or ascertaining whether anything is an instance of the term; (*ii*) the *intensional* connotation of a term consists in the range of terms or propositions implied by that term; (*iii*) the *objective* connotation of a term is comprised of the range of properties or characteristics common to the universe of entities the term denotes; (*iv*) the *subjective* connotation of a term consists in all the other terms associated with the term by the user or the criterion he might entertain respecting its usage.[2]

Viewed from the standpoint of this analysis of meaning, it becomes obvious that the conventional connotation of $_{S}H_{R}$ does not exhaust its full meaning, and that even while the term may function analytically in one context of inquiry, to be empirically significant it must be capable of functioning factually in other contexts. And at least in these other contexts, the term will not function as a synonym but rather factually, i.e., with denotative meaning.

If logico-mathematical signs and fictive terms exhaust the range of nonfactual or nondenotative terms, and if intervening variables are neither of these, they are either nonsense or else they *do* possess denotative significance. Were they nonsense, it is inconceivable that they could perform any role in scientific inquiry. The fact is that the attribution of a number implies a measurable parameter and only things, properties, and relations can be counted or measured. While some things which do not or cannot exist can be conceived as measurable or countable, we have already excluded fictions as the references of intervening variables. Accordingly, since MacCorquodale and Meehl admit that intervening variables can be measured, it follows that they do possess denotative meaning. And this conclusion is in line with the interpretation of intervening variables alluded to, namely, that they are terms whose definition is given by the laws in which they occur, which they imply, or which they presuppose.[3]

Are Intervening Variables Disposition Terms?

MacCorquodale and Meehl cite the concept of resistance as a physical instance of intervening variables thereby identifying them with dispositional concepts as characterized by Carnap (4). However, even while MacCorquodale and Meehl profess to mean exactly what Carnap means by dispositional concept,

[2] While roughly stated, this outline is sufficiently precise for our purposes here. The four senses of connotation are not unrelated. In many analyses of meaning, the conventional and intensional senses of connotation are combined under one rubric. Again, the subjective connotation may overlap with any of the others. It may be noted that a term's conventional connotation may be rendered by an operational definition. Clearly, these do not exhaust the meaning of a term (see 6).

[3] Bergmann (2) has also maintained, on somewhat different grounds, that intervening variables and hypothetical constructs cannot properly be distinguished on the issue of denotative meaning. However, he concludes that owing to this, they cannot be distinguished on any grounds whatsoever.

they depart from his analysis in crucial ways.

"Resistance," they argue, "is 'operational' in a very direct and primitive sense" (14, p. 96). In contrast to hypothetical constructs, the entire meaning of intervening variables is exhausted by the sentences about them or is reducible to sentences about impressions. The impression conveyed is that a set of reduction sentences wholly determines the meaning of an intervening variable once and for all. However, if intervening variables are dispositional concepts, this is not the case.

Carnap explicitly states that "a set of reduction pairs is a partial determination of meaning only and can therefore not be replaced by a definition. Only if we reach, by adding more and more reduction pairs, a stage in which all cases are determined, may we go over to the form of a definition" (4, pp. 450–451). And it is doubtful whether such a stage can ever be actually realized at least with all such terms. That is, particularly in the case of the more fertile dispositional terms in science, there is what Hempel calls an "openness of meaning" (7, p. 29) or an indeterminacy of application in respect to not-as-yet discovered possibilities.

In contrast to MacCorquodale and Meehl's supposition that intervening variables lack existential reference, Carnap conceived of dispositional concepts as a kind of predicate which, in the "thing-language," would mean that dispositional concepts refer to real or factual properties. According to Carnap and others who have analyzed the logical status of dispositional concepts (3, 5, 9), terms like resistance and $_SH_R$ refer to existences in a sense not radically different from such terms as "table" and "electron."

By disposition we mean to refer to such properties in the world as become differently manifest, in predictable ways, depending upon the complex of causal factors to which they are subjected. Corresponding to alterations in the causally effective context, the manifest aspect of the disposition may vary, whether in quality or quantity, within a determinate range of possibility. It is this range which comprises the potency of an entity so distinguished. In virtue of the potency of things, they both change and endure in intelligible fashion. But in few if any cases can we rationally predict all the causal factors which might conceivably influence the manifest aspect or appearance character of a dispositional property. There is little doubt that $_SH_R$ denotes a dispositional property. So far as it is a property of organisms to which we refer, the question of existence is not simply confined to whether the equation for acquisition holds. Whether any organism is capable of possessing any kind or degree of $_SH_R$ is a factual issue and thus one which is settled in terms of appropriate empirical evidence. Therefore, the question of whether $_SH_R$ exists or not is no more trivial than any other such question posed in respect to any other entity or property which can be named. The introduction of an intervening variable is thus never merely a matter of convenience even though our convenience is served by such introduction.

ARE INTERVENING VARIABLES SIMPLY CONVENTIONAL?

In their summary, MacCorquodale and Meehl conclude that "the only rule for proper intervening variables is that of convenience, since they have no factual content surplus to the empirical functions they serve to summarize" (14, p. 107). Elsewhere they maintain that apparently legitimate sets of intervening variables could be obtained by "various arbitrary groupings and combinations" of the relevant equations (p. 98).

It is generally held that no concept in science is to be regarded as factually significant or as denoting a state of affairs in the world unless it has systematic purport, unless, that is, it is involved in other principles besides that for which it was invented or introduced. *Ad hoc* concepts either undergo such enhancement of purport or else they are counted spurious or empty and are discarded. Because such enrichment in the meaning of any concept must always be expected, any conceptual analysis must provide for this possibility in allowing that no concept can receive a final and full explication on the basis of its present functions in inquiry or the existing evidence by which it is supported. This openness of meaning implies that there is always an indeterminate range of factual content implicit in any significant concept.

A brief account of some scientific history will illustrate this point. At the beginning of this century, Planck, concerned to explain certain discrepancies between classical theory and experimentally observed facts of incandescence and radiation, postulated a new universal constant of nature which he called *h*. The principles containing this constant imply that emission and absorption processes both occur discontinuously, being in the nature of jumps of finite magnitude. Although heralding a great revolution in scientific thought, a prominent reason why Planck's theory initially met with a skeptical reception and could not rightly be accepted with studied confidence was the limited domain of applicability of *h*, or its poverty of factual bearings.

However, a few years later Einstein formulated a law for the kinetic energy of an ejected electron in the photoelectric phenomenon which required *h* as a constant of proportionality. Since then, *h* has been employed in a variety of contexts and now not only furnishes the basis for explaining the intensity of radiation and the wave length for which it represents a maximum (and for which it was originally formulated), but also for interpreting the quantitative relations existing in many other cases, namely, to cite a few, the specific heat of solids, the photochemical effects of light, the orbits of electrons in the atom, the wave lengths of the lines of the spectrum, the frequency of the X rays produced by the impact of electrons of given velocities, the velocity with which gas molecules can rotate, the distances between the particles which make up a crystal, and others.

In short, concepts which are formulated to explain a given limited subject matter may prove applicable in other spheres and it is this which certifies their significance or systematic utility (cf. 7, pp. 52–54). In the case of Planck's quantum theory, the relevance of *h* in a variety of situations means, in the words of D'Abro, that the theory "cannot be regarded as a mere makeshift devised for the sole purpose of interpreting the law of equilibrium radiation. We must recognize therefore that the quantum theory has uncovered a new world of physical occurrences, a world formerly unsuspected" (1, p. 471).

Now it might be a permissible convention to hold that so far as a concept applies only to the area for which it was formulated it is an intervening variable, but as it becomes relevant to other, previously independent, subject matters, it acquires the status of a hypothetical construct. However, it is not clear what value would come of this convention, especially since it would present problems such as fixing the point at which the conversion may be said to have occurred.

In similar vein, it often happens that a concept defined in terms of a certain law is rendered more precise and general in being amplified by a higher-level law or theory. The concept of resistance, conceived as an intervening variable by MacCorquodale and Meehl, is a case in point. Defined in terms of Ohm's law, resistance is a constant whose value is the ratio of voltage and current, a linear function. The term denotes that property of conductors which determines the amount of current which will flow when a given amount of voltage is impressed. However, it turns out that Ohm's law is inadequate in the case of such conductors as gas tubes, where, if the voltage reaches beyond or below a certain range, resistance no longer remains a constant, i.e., the slope of the curve gradually flattens. Here resistance becomes a function of other factors not provided for in Ohm's law, e.g., the field effects of electron clouds.

It is not the case, then, that Ohm's law exhausts the meaning of "resistance" nor is it the case that in being incorporated within higher-level laws, the term acquires denotative properties which it previously lacked entirely. It must always be realized that with new discoveries or new theoretical formulations, the meaning of any concept may be revised, refined, or broadened in being incorporated within new relationships. And it is precisely this application to and corroboration by diverse, independent domains that reinforces our confidence in a concept's factual significance.

The concept $_{S}H_{R}$ appears to fulfill these requirements since it has been applied to such prima facie diverse areas as the learning of nonsense syllables, mazes, etc., and the development of personality and neurotic symptoms, and to all the high-level mental functions of humans. In fact, it is in virtue of the capacity of $_{S}H_{R}$ to be differently specialized in different areas of application that may make Hull's system qualify as a theory in the sense outlined.[4] Accordingly, postulate 4 does not exhaust the meaning of $_{S}H_{R}$, and, if anything, the term denotes a dispositional property of organisms and this is something real or true about certain organisms and, in turn, the world.

If the foregoing arguments are correct, it follows that so far as intervening variables are to be counted as factually significant, they are not merely names for arbitrary combinations of empirical equations. In the first place, no empirical equation can be manipulated in every logically possible way and be expected to yield statements, all of which are empirically significant or scientifically useful. Many legitimate logical or mathematical operations are proscribed because the results would be absurd or completely useless. For example, in multiple-factor analysis, it is logically possible to rotate the axis in an infinite number of ways, yet, aside from the logical requirement that uniqueness and simple structure be attained, there is the empirical requirement that the rotation accord with theoretical or factual considerations. Again, the curves by which a functional law might be represented graphically are rarely (if ever) extended in all possible directions nor are all the branches exploited. The reason is that such extrapolations would either have no empirical significance or else would contravene fact or other principles.

[4] Whether Hull's system actually is a theory in the sense outlined is problematic. First, many of the formulae are derived from curve fitting, and second, it is far from being the kind of rigorous deductive system as are the physical theories cited. The fact that $_{S}H_{R}$ can be defined in experimental terms (as in the formula for acquisition) is not atypical or contrary to the nature of theories. The essential point is whether it can be fully or exhaustively defined in this way. If so, Hull's system would likely qualify only as a system of laws.

Furthermore, in constructing theories, economy of postulation is the most reliable insurance against inconsistency. We mean by theoretical simplicity that the number of independent factors, properties, principles, or variables be held to a minimum compatible with the scope or level of penetration desired of the theory. A new concept may be introduced into a system only if it does not produce redundant conclusions or result in inconsistencies, and only where its inclusion will enhance the predictive power of the theory. Whether MacCorquodale and Meehl's suggested concept of cumulative reinforcement would fulfill these requirements and the others noted previously is not demonstrated by them (although it appears to coincide with recent restatements of his position by Hull [11]). At any rate, if this concept is to be included within Hullian theory, it would not be an arbitrary matter or simply one of convenience. And if included, it would denote something about organisms, something which exists, rather than being simply a name for an equation and thus referring to nothing but itself. In brief, there are always methodological and empirical considerations which proscribe attaching factual significance to all possible transformations of equations or groupings of variables. Logical validity is not identical with factual truth or significance.

One last historical point is in order. The big difficulty in establishing the conception of blood circulation was the absence of any visible connections between the terminal arteries and the veins. Harvey postulated the existence of capillaries, but it was only after the microscope was considerably improved that these were actually observed as real. In similar manner, to explain his experimental results, Mendel postulated hereditary units which only recently have been microscopically and chemically identified as genes or real entities. Even while the advent of new confirmatory evidence increased the probability of these hypotheses, it did not result in the conversion of something lacking "a surplus meaning which is existential" into something possessing such "meaning." The relation between evidence and a hypothesis is logical and not predicative.

VARIABLES

If our criticisms are correct and justified, much of the rationale behind MacCorquodale and Meehl's distinction between intervening variables and hypothetical constructs is untenable. While upholding the distinction itself, a sounder rationale is suggested by the logical differences between empirical laws and theories. An intervening variable would refer to any concept whose definition is provided by laws or a set of laws whereas a hypothetical construct would refer to any concept whose definition is provided within a theoretical system. Intervening variables or law-like concepts would thus be distinguished in being introduced by experimental or operational procedures whereas hypothetical constructs or theory-like concepts would be distinguished in being introduced by postulational procedures, e.g., by interpreting the primitive apparatus of a formal calculus. Because they are more intimately associated with observation terms, intervening variables are introduced either via explicit or conditional definitions based upon invariable associations, functional relations, or causal involvements obtaining among entities or processes, among these and their properties, or among the properties themselves. Explicit definitions are usually of the form $X = \text{def.}\ Y$ often considered nominal or analytic. Conditional definitions involving reduction chains are necessary for introducing dispositional terms and take the form if X is at p, then X is q if and only if X is r (see 4 and 7 for details).

The differences in logical status between intervening variables and hypothetical constructs are therefore based upon the differences between laws and theories. Accordingly, whatever characteristics distinguish laws and theories attach to intervening variables and hypothetical constructs as well. Many of MacCorquodale and Meehl's remarks fit in with this rationale.

Understood in this way, the relations between intervening variables or hypothetical constructs on the one hand, and independent or dependent variables on the other, become a function of problematic context. Rather than describing an absolute logical or meaning property of the terms so characterized, the distinction reduces to a matter of how equations are expressed or what facet of the inclusive psychological state of affairs is being inquired into or projected as problematic. To understand how a sensory complex issues in a response may require reference to intervening processes and any law adumbrated for this purpose may consist in three differentiable components, an independent variable, an intervening variable, and a dependent variable. But what functions as an independent variable in one context may function as an intervening variable or a dependent variable in another. If it is the sensory complex as such that we are concerned with, then the stimulus object may constitute the independent variable and the sensory complex the dependent variable. What might, under some circumstances, constitute an intervening variable, e.g., a neural process, may be an independent or dependent variable in, e.g., a physiologically oriented inquiry.

In general, intervening variables (whether these be law-like or theory-like) are introduced for either of two reasons: (*a*) without such an introduction, explanation may be insuperably partial or prediction inexact or (*b*) without such an introduction, explanation may be extremely cumbersome and require a host of disjoined *ad hoc* adjustments to be kept up to date. Thus, to achieve systematic elegance or a deeper and more precise understanding of subject matter, the introduction of intervening variables or hypothetical constructs may be essential. The test for their appropriateness or correctness has been stated quite often. In essence it consists in (*a*) the possibility, not otherwise realizable, of generating new and ever more precise conclusions out of a set of laws or a theory and experimentally corroborating these and (*b*) demonstrating a logical coherence among the principles or concepts of the system. It is these requirements which Hull has often described as the secure anchoring of intervening variables on both sides to observable and measurable conditions or events (10, p. 22). This "anchoring" is both a logical and an empirical or evidential affair. Incidentally, Hull's choice of the "anchor" metaphor is most apt: if the anchor be taken as observation and mensuration sentences, and the ship as the theory, it follows that the two are not identical or interchangeable and that there are different degrees of "security."

One last point needs to be made. In contrast to the position maintained here, certain writers, e.g., Kendler (12) and Marx (15), have either repudiated the distinction between intervening variables and hypothetical constructs or deprived it of logical significance. Like other orthodox operationists, they have offered a monolithic point of view in its stead. According to them, only intervening variables are to be encouraged or permitted into the corpus of mature science since hypothetical constructs are regarded either as metaphysical hypostatizations and thus "meaningless" or as transitional crutches at best. In fixing upon certain of MacCorquodale and Meehl's arguments concerning the supposed purely conventional, denotative-less status of intervening variables, Kendler and Marx have ignored or misread the valid arguments raised in support of hypothetical constructs as denoting. One reason for this misreading may stem from the lack of clarity in MacCorquodale and Meehl's paper, but the main reason must be attributed to a severe nominalistic bias. At any rate, the whole tenor of this paper and the one previously referred to (6) renders a more detailed examination of Kendler's and Marx's views unnecessary.

Summary

Two basic questions were raised: (*a*) Are there significant logical differences between intervening variables and hypothetical constructs? (*b*) Are these correctly conceived by MacCorquodale and Meehl? The reply to the first question was made by pointing out that the differences relate to those distinguishing empirical laws from theories. The reply to the second question was made by pointing out certain ambiguities and misconceptions in MacCorquodale and Meehl's thesis, the most prominent of which was the assumption that intervening variables were simply conventional and without denotative meaning. During the course of this critical analysis, the senses or modes of meaning, the nature of disposition concepts, the function and character of abstract terms, and the kinds and functions of variables were touched upon.

References

1. D'Abro, A. *The decline of mechanism.* New York: Van Nostrand, 1939.
2. Bergmann, G. Theoretical psychology. *Annu. Rev. Psychol.*, 1953, 4, 435–458.
3. Broad, C. D. The "nature" of a continuant. In H. Feigl & W. Sellers (Eds.), *Readings in philosophical analysis.* New York: Appleton-Century-Crofts, 1949. Pp. 472–481.
4. Carnap, R. Testability and meaning. *Phil. Sci.*, 1936, 3, 420–471; 1937, 4, 1–40.
5. Chisholm, R. M. The contrary-to-fact conditional. In H. Feigl & W. Sellars (Eds.), *Readings in philosophical analysis.* New York: Appleton-Century-Crofts, 1949. Pp. 482–497.
6. Ginsberg, A. Operational definitions and theories. *J. gen. Psychol.*, in press.
7. Hempel, C. G. Fundamentals of concept formation in empirical science. *Int. Encyc. Unif. Sci.*, II, 7.
8. Hilgard, E. R. *Theories of learning.* New York: Appleton-Century-Crofts, 1948.
9. Hofstadter, A. Professor Ryle's category-mistake. *J. Phil.*, 1951, 48, 257–270.
10. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
11. Hull, C. L. *Essentials of behavior.* New Haven: Yale Univer. Press, 1951.
12. Kendler, H. H. "What is learned?"—A theoretical blind alley. *Psychol. Rev.*, 1952, 59, 269–277.
13. Lewis, C. I. *An analysis of knowledge and valuation.* La Salle, Ill.: Open Court, 1946.
14. MacCorquodale, K., & Meehl, P. E. On a distinction between hypothetical constructs and intervening variables. *Psychol. Rev.*, 1948, 55, 95–107.
15. Marx, M. H. Intervening variable or hypothetical construct? *Psychol. Rev.*, 1951, 58, 235–247.
16. Skinner, B. F. Are theories of learning necessary? *Psychol. Rev.*, 1950, 57, 193–216.
17. Tolman, E. C. A psychological model. In T. Parsons & E. A. Shils (Eds.), *Toward a general theory of action.* Cambridge: Harvard Univer. Press, 1952. Pp. 279–361.

(Received May 18, 1953)

Psychological Review
Vol. 61, No. 2, 1954

SOME VIEWS ON MATHEMATICAL MODELS AND MEASUREMENT THEORY[1]

C. H. COOMBS, H. RAIFFA,[2] AND R. M. THRALL

University of Michigan

We shall undertake first to review the role of mathematical models in a science and then briefly discuss the models used in classical measurement theory. This will be followed by a generalization of measurement models. Illustrations will be introduced when needed to clarify the concepts discussed.

The Role of Mathematical Models

We shall use the terms *physical objects*, *real world*, and *object system* synonymously to signify that which the empirical scientist seeks to study, including such objects as opinions or psychological reactions. The scope and content of a domain is selected by the scientist with the intent of discovering laws which govern it, of making predictions about it, or of controlling or at least influencing it.

There are potentially at least as many ways of dividing up the world into object systems as there are scientists to undertake the task. Just as there is a potential variety of object systems, so also is there a potential variety of mathematical systems. Let us describe the nature of a mathematical system:[3] A mathematical system consists of a set of assertions from which consequences are derived by mathematical (logical) argument. The assertions are referred to as the axioms or postulates of a mathematical system. They always contain one or more primitive terms which are undefined and have no meaning *in the mathematical system*. The axioms of the mathematical system will usually consist of statements about the existence of a set of elements, relations on the elements, properties of the relations, operations on the elements, and the properties of the operations. Particular mathematical systems differ in the particular postulates which form their bases. It is evident then that the variety of mathematical systems is limited only by the ability of man to construct them.

Our view of the role that mathematical models play in a science is illustrated in Fig. 1. With some segment of the real world as his starting point, the scientist, by means of a process we shall call abstraction (A), maps his object system into one of the mathematical systems or models. By mathematical argument (M) certain mathematical conclusions are arrived at as necessary (logical) consequences of the postulates of the system. The mathematical conclusions are then

[1] This paper is a chapter in a forthcoming book, *Decision Processes*, edited by C. H. Coombs, R. L. Davis, and R. M. Thrall, published by John Wiley and Sons, Inc. It is an outgrowth of a number of sessions on measurement theory in an interdisciplinary seminar on the application of mathematics to the social sciences held during the academic year 1951–1952 by Drs. Coombs, Raiffa, and Thrall. While the influences of the separate authors are sometimes distinguishable, the paper is a truly joint product. The authors are listed alphabetically. This research was carried out under Office of Naval Research Contract Nonr 374 (00) NR 041–011 and a Ford Foundation Behavioral Studies grant.

[2] Dr. Raiffa is now at Columbia University.

[3] For a more detailed discussion of the nature of mathematical systems, see (8, 12, and 13).

converted into physical conclusions by a process we shall call interpretation (I).

Let us start with a specific real-world situation $(RW)_1$ and by process A map it into a mathematical system $(MS)_1$. We can look at $(MS)_1$ as a *model* of $(RW)_1$. Looked at in reverse, we can start with consideration of $(MS)_1$ and then $(RW)_1$ can be viewed as a *model* of $(MS)_1$, and the process of going from $(MS)_1$ to $(RW)_1$ we call "realization." Thus, "realization" is the converse of "abstraction." Now, given $(MS)_1$ we might be able to find a real-world situation $(RW)_2$ such that by assigning meanings to the undefined terms of the mathematical system, the assertions about "sets of elements," "relations," and "operations" in $(MS)_1$ become identified with objects or concepts about $(RW)_2$. That is, $(RW)_2$ may be another model of $(MS)_1$, and the process of going from $(RW)_1$ to $(MS)_1$ to $(RW)_2$ often indicates subtle analogies between systems such as $(RW)_1$ and $(RW)_2$. To the mathematician who often starts with an abstract system the model is a concrete analogue of the abstract system. To the social scientist who starts with phenomena in the real world the model is the analogue in the abstract system.

In establishing a model for a given object system one of the most difficult tasks is to attempt a division of the phenomenon into two parts; namely that part which we abstract (A) into the basic assumptions or axioms of the abstract system, and that part which we relegate to the physical conclusions and which we reserve as a check against the interpretations from the abstract system. In a given object system there is no unique partition of the phenomena, and which partition is made depends on the creative imagination of the model builder. Indeed, there are models in the physical and biological sciences for which there are no experimentally verified or verifiable correlates in the real world for the undefined terms, relations, and operations in the abstract model. A similar situation prevails on the side of the abstract system; i.e., it is often possible in a given abstract system to interchange the roles of certain axioms and theorems. Thus, in a given system there is no unique method of splitting the mathematical propositions into axioms and theorems. In going from the abstract system to the object system we have the parallel processes of realization and interpretation. It is quite common to consider these synonymous; however, we prefer in this discussion to reserve the word "interpretation" for the process which maps the mathematical conclusions (rather than the axioms) into the object system.

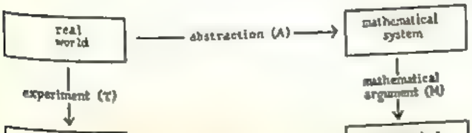

FIG. 1. The symmetrical roles of experiment and mathematics

Let us summarize briefly up to this point. Beginning with a segment of the real world, the scientist, by an entirely theoretical route, has arrived at certain conclusions about the real world. His first step is a process of abstraction from the real world, then a process of logical argument to an abstract conclusion, then a return to the real world by a process of interpretations yielding conclusions with physical meaning. But there is an alternative route to physical conclusions and this is by way of working with the object system itself. Thus, the scientist may begin with the real-world segment in which he is interested and proceed directly to physical conclusions by a process of observation or experiment (T).

The path (T) (experimentation) from the real world to the physical conclusions needs further scrutiny. Usually in theory construction the scientist embarks on model building after he has many facts at his disposal. These facts he partitions into two parts—one part serves as a springboard for the abstraction process (A); the other part serves as a check on the model by making comparisons with these initial facts and the interpretations (I) stemming from the model. If a specific interpretation is not at variance with a fact in the initial reservoir, but at the same time not corroborated by our a priori notions of the object system, then the model *perhaps* has contributed to our knowledge of the object system. The scientist next tests this tentative conclusion by setting up a plan of experimental verification, if this is possible. Often direct verification may not be possible, and corroboration stems from examination of experimental evidence which supports claims of the model quite indirectly. That is, motivated by interpretations of the model, the scientist sets up an experimental design, obtains observations by experimentation, makes a statistical interpretation of these observations into physical conclusions, and compares the conclusions with those of the abstract route in order to appraise the model. As suggested by Frederick Mosteller[4] of Harvard University, it would be appropriate to generalize Fig. 1 as shown in Fig. 2. The route A_2EI_2 in Fig. 2 is summarized by the route T in Fig. 1. If the physical conclusions of the process A_1MI_1 are at variance with the a priori facts or with conclusions arrived at via A_2EI_2 (and if more confidence is placed in the experimental route than in the theoretical route), then the suitability of the model is suspect.

The task of a science looked at in this way may be seen to be the task of trying to arrive at the same conclusions about the real world by two different routes: one is by experiment and the other by logical argument; these correspond, respectively, to the

[4] Personal communication.

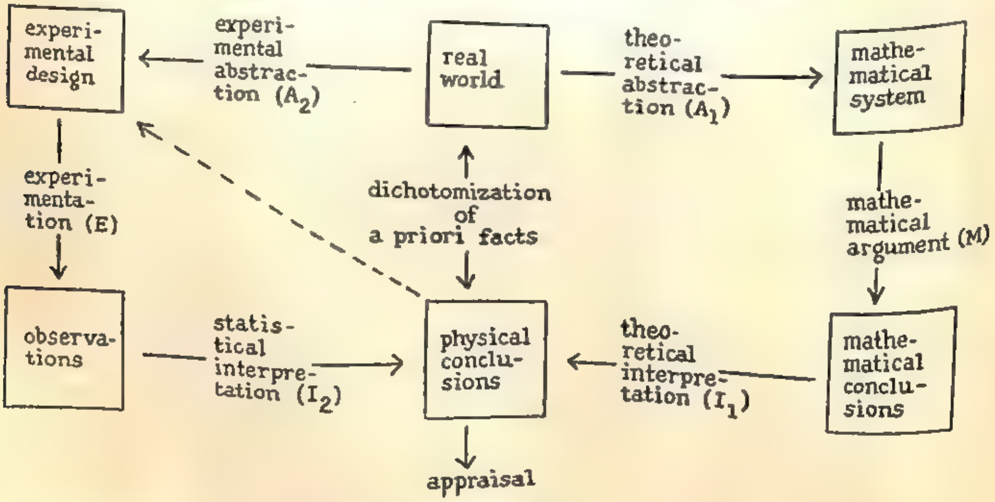
experimental design
experimental abstraction (A_2)
real world
theoretical abstraction (A_1)
mathematical system
experimentation (E)
dichotomization of a priori facts
mathematical argument (M)
observations
statistical interpretation (I_2)
physical conclusions
theoretical interpretation (I_1)
mathematical conclusions
appraisal

Fig. 2. A generalization of Fig. 1

left and right sides of Fig. 1 and 2. There is no natural or necessary order in which these routes should be followed. The history of science is replete with instances in which physical experiments have suggested axiom systems to the mathematicians and, thereby, contributed to the development of mathematics. On the other hand, mathematical systems developed under such stimulation, in turn, have suggested experiments. And there have been many instances of mathematical systems developed without reference to any known reality which subsequently filled a need of theoretical scientists. The direction that mathematics has taken is in considerable part due to its interaction with the physical sciences and the problems arising therein.

It is illuminating here to observe the way in which the models of the mathematical theory of probability and statistics fit this picture. As in any abstract system, the mathematical theory of probability is devoid of any real-world content; and as in any other mathematical system, the axioms of probability specify interrelationships among undefined terms. It is common to let the notion of probability itself be undefined and to attempt to capture in the axiomatic structure properties of probability motivated by the interpretations we have in mind (e.g., gambling games, processes of physical diffusion, etc.). Given an association of probabilities to prescribed elementary sets, the axioms of probability dictate how one must associate probabilities with other sets. How we make these preliminary associations, provided that we have consistency, is not relevant to the purely abstract system. When we come to apply the probability model, we are confronted with the problem of identifying real events with abstract sets in the mathematical system and the measurement problem of associating probabilities to these abstract sets. Experience has taught us that if we exploit the notion of the relative frequency of occurrence of real events when making our preliminary associations, then the interpretations from the model have a similar frequency interpretation in the real world. To be sure, our rules of composition in the formal system were devised with this in mind. We associate probabilities in one way rather than another in the process A so that when we generate AMI, our interpretations are in "close" accord with results of experimentation, T, when T is possible. When T is not possible we have to rely to a great extent on analogy.

An extremely important problem of statistics can be viewed as follows: For a priori reasons we may have a well-defined family of possible probability associations. Each element of this family, when used in the abstraction process A, generates by AMI a probability measure having a frequency interpretation over real events. In addition, we are given a set of possible actions to be taken. Preferences for these actions depend in some way on the relative "appropriateness" of different probability associations in the abstraction process, A. By conducting an experiment, T, and noting its outcome we gain some insight into the relative "appropriateness" of the different probability associations and thus base our action accordingly. Variations of this problem, which involves the entire AMI-T process, have been abstracted sufficiently so that models of mathematical statistics include counterparts of all these ingredients within the mathematical system itself.

In a given model we may be confronted with the problem of deciding whether the *AMI* argument gives results "close enough" to the experimental results from T. We often can view this problem involving a complete *AMI-T* process as the real-world phenomenon to which we apply the A process, sending it into a formal mathematical statistics system. The statistics system analyzes step M, and our interpretation I takes the form of a statement of acceptance or rejection concerning the original theory.

The process of measurement, corresponding to A in Fig. 1, provides an excellent illustration of the role of mathematical models. There are many types of observations that can be called "measurement." Perhaps the most obvious are those made with yardsticks, thermometers, and other instruments, which result immediately in the assignment of a real number to the object being measured. In other cases, such as the number of correct items on a mental test or the size of a herd of cattle, the result of measurement is a natural number (positive integer). In still other cases, such as relative ability of two chess players, relative desirability of a pair of pictures, or relative hardness of two substances, the result is a dominance (or preference) relation. We might even stretch the concept of measurement to include such processes as *naming* each element of some class of objects, or the photographic representation of some event, or the categorization of mental illnesses or occupations.

The process of measurement may be described formally as follows. Let $P = \{p_1, p_2, \cdots\}$ denote a set of physical objects or events. By a measurement A on P we mean a function which assigns to each element p of P an element $b = A(p)$ in some mathematical system $B = \{b_1, \cdots\}$. That is, to each element of P, we associate an element of some abstract system B (the process A of Fig. 1). The system B consists of a set of elements with some mathematical structure imposed on its elements. The actual mapping into the abstract space B comprises the operation of measurement. The mathematical structure of the system B belongs to the formal side of measurement theory. The structure of B is dictated by a set of rules or axioms which states relationships between the elements of B. However, no connotation can be given to these elements of B unless it is explicitly stated in the axioms; i.e., their labels are extraneous with respect to considerations of the structure of B.

After making the mapping from P into B, then one may operate with the image elements in B (always abiding by the axioms, process M of Fig. 1). Purely mathematical results obtained in B must then be interpreted back in the real world (the process I of Fig. 1) to enable one to make predictions or to synthesize data concerning set P.

If the manifestations of P (as a result of the process T of Fig. 1) are in conflict with the results of process I obtained from B, then one must search for a new cycle *AMI*. Suppose that we have a family of abstractions $\{A_\alpha\}$ from the given situation P, and suppose that M_α, I_α complete the cycle begun with A_α. Among all of the available cycles $A_\alpha M_\alpha I_\alpha$ we seek one, say $A_0 M_0 I_0$, which is "closest" to T according to some criterion. Some models have a criterion built in to judge closeness, and others of a more deterministic nature require an exact fit.

The process T represents the experimental or operational part of model building; the process M represents the formal or logical aspect. The processes A and I are really the keys to the model and serve as bridges between experiment and formal reasoning.

It might be well here to draw clearly the distinction between a model and a theory. A model is not itself a theory; it is only an available or possible or potential theory until a segment of the real world has been mapped into it. Then the model becomes a theory about the real world. As a theory, it can be accepted or rejected on the basis of how well it works. As a model, it can be right or wrong only on logical grounds. A model must satisfy only internal criteria; a theory must satisfy external criteria as well.

An example of the distinction between models and theories lies in the domain of measurement. A measurement scale, such as an ordinal, interval, or ratio scale, is a model and needs only to be internally consistent. As soon as behavior or data are "measured" by being mapped into one of these scales, then the model becomes a theory about those data and may be right or wrong. Scales of measurement are only a very small portion of the many formal systems in mathematics which might serve as image spaces or models, but will be discussed here as they constitute very simple and immediate examples of the role of mathematical models. First to be discussed will be the models of conventional measurement theory, and then a generalization of these models will be presented.

Mathematical Models of Classical Measurement Theory

The first comprehensive classification of the mathematical models used in conventional measurement theory was made by Stevens (9). He classified scales of measurement into nominal, ordinal, interval, and ratio scales, the latter two christened by him. A more complete discussion of these scales is contained in a later work by him (10), and also in Coombs (3, 5) and Weitzenhoffer (11). Because of the available literature on these scales and because they constitute a restricted class of models, they will be briefly summarized here only to provide a basis for generalization in the next section.

The mathematical model of measurement is said to be *nominal* if it merely contributes a mapping A_0 of P into M_0 without imposing any further structure on M_0. A nominal scale M may be subjected to any 1–1 transformation without gain or loss in information.

An *ordinal* scale of measurement is implied if there is a natural ranking of the objects of measurement according to some attribute. More precisely, the ordinal scale is appropriate if the objects of measurement can be partitioned into classes in such a manner that (*a*) elements which belong to the same class can be considered equivalent relative to the attribute in question; (*b*) a comparative judgment or an order relation can be made between each pair of distinct classes (for example, class x is more ____ than class y); (*c*) there is an element of consistency in these comparative judgments —namely, if class x is more ____ than class y and class y is more ____ than class z, then class x is more ____ than class z (that is, the comparative judgment or order relation is transitive). For example, the familiar socioeconomic classes, upper-upper, lower-upper, upper-middle, lower-middle, upper-lower, and lower-lower, imply the measurement of socioeconomic status on an ordinal scale. The numbers 1, 2, 3, 4, 5, 6, or 1, 5, 10, 11, 12, 14, or the letters A, B, C, D, E, F could designate the six classes without gain or loss of information.

The measurement is said to be an *interval* scale when the set M consists of the real numbers and any linear transformation, $y = ax + b$ $(a \neq 0)$, on M is permissible. Measurement on an interval scale is achieved with a constant unit of measurement and an arbitrary zero. An example of an interval scale is the measure of time. That is, "physical events" can be mapped into the real numbers and all the operations of arithmetic are permissible on the differences between all pairs of these numbers.

If the set M consists of the real numbers subject only to the transformation group $y = cx$ where c is any nonzero scalar, the scale is called a *ratio* scale. Measurement on a ratio scale is achieved with an absolute zero and a constant unit of measurement. The scalar c signifies that only the unit of measurement is arbitrary. In a ratio scale all the operations of arithmetic are permissible. The most familiar examples of ratio scales are observed in physics in such measurements as length, weight, and absolute temperature.

A GENERALIZATION OF MEASUREMENT MODELS

An axiomatic basis for certain scales of measurement will be presented in this section. Other scales can be generated by forming mixtures (or *composites*) of these. Indeed, some of the scales listed in the diagram shown in Fig. 3 can be regarded as composites of others.

We will now list defining axioms for each of these systems and briefly discuss their roles. It is not claimed that this list is exhaustive; it is presented to illustrate certain possibilities for significant generalizations of scales used in the classical theory. The arrangement in the diagram is from top to bottom in order of increasing strength of axioms; a connecting line indicates that the lower listed system is a special case of the higher one.

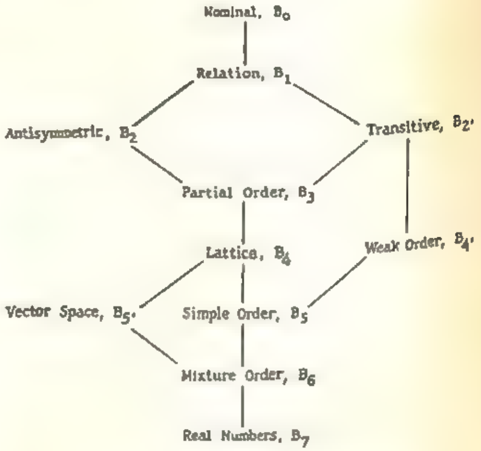

FIG. 3. Measurement scales

B_0, *the nominal scale.* A nominal scale, B_0, may be considered a mathematical system consisting merely of a set of elements. We define the *index* of B_0 to be the number of elements in B_0. (The index may be finite or infinite.)

Examples of segments of the real world that are mapped into nominal scales are psychiatric classifications, job families, and disease types.

The nominal scale, B_0, is the most primitive step in any system of measurement. The process of naming partitions a set into classes such that there is a relation of "equality" or equivalence between pairs of elements from the same class. The nominal scale is fundamental since the process of discrimination is a necessary prerequisite for any more complex form of measurement.

B_1, *the relation scale.* Perhaps the smallest step that may be taken to strengthen this mathematical system is to introduce a relation between some pairs of elements. In technical language, a relation R on a set B_1 is a set of ordered pairs (b, b') of elements of B_1. We write bRb' to indicate that (b, b') is one of the pairs included in the relation R, and call the set B_1 a relation scale. It is important to recognize that for R to constitute a very useful relation, not all possible pairs (b, b') from B_1 can be included in the relation R.

With some risk of misinterpretation or distortion, these concepts might be illustrated as follows. Consider a set of persons identified by a nominal scale, B_0. Let us now define the relation R on B_0 to be "loves." Thus R consists of the ordered pairs (a, b) for which, a loves b.

The particular relation used here as an illustration is one whose mathematical properties are mostly negative. We cannot conclude from a loves b and b loves c that a loves c, or that b loves a, or that b does not love a. For example, if John loves Mary and if Mary loves Peter, it may well be that, far from loving him, John would like to see Peter transported to the South Pole. In the terminology to be introduced below, we would say that love is not symmetric, is not asymmetric, and is not transitive.

B_2, *the antisymmetric relation scale.* A relation R on a set B is said to be *antisymmetric* if aRb and bRa together imply that a is identical with b. An example is the relation $\geqq$ for real numbers. A statement such as "picture a is at least as good as picture b" illustrates an antisymmetric relation on a collection of pictures, provided that there are not in the collection two distinct pictures of equal merit, i.e., two pictures about which the judge is *indifferent.*

Closely connected to the concept of antisymmetry is that of asymmetry. A relation R on a set B is said to be *asymmetric* if aRb implies $bR'a$ (where $bR'a$ means that b is *not* in the relation R to a). The mathematical prototype of asymmetry is the relation $>$ for real numbers. Verbal forms for asymmetric relations include such statements as "picture a is better than picture b," or "player a beats player b in a game."

Antisymmetry and asymmetry are seen to be at the root of statements of comparison. These two classes of relations can be regarded as the most primitive types of order relations. At the opposite pole from these concepts is that of symmetry. A relation R is said to be *symmetric* if aRb implies bRa. For example, the relations "is a sibling of," "is a cousin of," and "is the same color as" are all symmetric.

If S is an asymmetric relation on a set B, we can obtain from it an antisymmetric relation R by the definition: aRb means either aSb or $a = b$. Conversely, if R is antisymmetric and we define aSb to mean aRb and $a \neq b$, then S is asymmetric. It is customary to use the symbols $\leqq$, $\geqq$ for antisymmetric relations and to use $<$, $>$ for the associated asymmetric relations.

$B_{2'}$, *the transitive relation scale.* A relation R is said to be transitive if aRb and bRc imply aRc. In the physical world, preference judgments which are not transitive are frequently regarded as inconsistent or irrational. However, situations such as that of three chess players, each of whom can beat one of the other two, show that transitivity is not a requirement of nature.

The chess player relation is antisymmetric but not transitive. An example of a relation that is symmetric and transitive is given by a communication system where each link is bidirectional; here aRb is given the meaning "there exists a chain of links starting with a and ending with b." If the links are not required to be bidirectional, the relation is still transitive but is no longer symmetric. Note that in this example it is quite possible to have aRa, i.e., a chain beginning at a and ending at a. (This chain must have at least one element different from a.)

The relation "a is the rival of b" (say as suitors of a particular girl) is symmetric and is almost transitive. If aRb and bRc, we can conclude aRc unless $a = c$; but we can hardly regard a as being his own rival. This type of relation arises frequently in studies of social structures. We say a relation R is *quasi-transitive* if aRb, bRc, and $a \neq c$ imply aRc. The sibling relation is also quasi-transitive. Of course, if R is quasi-transitive we can define a new relation S to be the same as R except that also aRb, bRa imply aSa. In some instances S is just as good a model as R, but in others the extension from R to S destroys the usefulness of the model.

As an example consider the structure matrix $A = \|a_{ij}\|$ of some society. Thus we set $a_{ij} = 1$ if person i has direct influence on person j and set $a_{ij} = 0$ otherwise. One must decide in accordance with the purpose of the investigation whether or not to set the diagonal element a_{ii} equal to 0 or to 1. (The relation "i has direct influence on j" is not transitive even if we take each $a_{ii} = 1$, but this example nevertheless illustrates the kind of problem involved in the contrast between transitivity and quasi-transitivity.)

If the relation aRb meant a "is higher in socioeconomic status than" b, and this required that a had more income *and* more education than b, then the relation R would be asymmetric and transitive.

B$_3$, *the partly ordered scale.* A relation $\geqq$ which is reflexive,[5] antisymmetric, and transitive is called a partial order. If for some pair (a, b) neither of the relations $a \geqq b$ nor $b \geqq a$ holds, we say that a and b are *incomparable* relative to $\geqq$. In the case of a preference relation, incomparability is not the same thing as *indifference*. We call a set a *poset* (partly ordered set) if there is a partial order relation defined on B.

If $a \geqq b$, we also write $b \leqq a$; if $a \geqq b$ and $a \neq b$, we write $a > b$ or $b < a$.

A partial order may be illustrated as follows. Suppose that on a mental test no two individuals in a group pass exactly the same items. Now let $a \geqq b$ symbolize the relation "a passed every item b did and perhaps more." Then $a > b$ means that "a passed all the items b did and at least one more." This poset reflects multidimensionality of the attributes mediating the test performance, and some interesting mathematical problems arise regarding the partial order as a "product" of simple orders. The result is a nonmetric form of factor analysis with some of the same problems as factor analysis (3).

Next we consider the mental test example modified so as to allow the possibility that two individuals a and b pass exactly the same items. Then in the above notation we have $a \geqq b$ and $b \geqq a$, but not $b = a$. Hence, $\geqq$ no longer gives a partial order. However, if we define $a \mathfrak{I} b$ to mean $a \geqq b$ and $b \geqq a$, it is not hard to show that if we identify individuals with the same test performance then $\geqq$ is a partial order relation. Or, alternatively, we could consider $\geqq$ as a partial order relation on the set of possible test performances. It is customary to make such identifications and speak of a partial order as if it were actually on the initial set rather than on the identified classes or on the test results.

[5] A relation is reflexive if it holds between an element and the same element, symbolized aRa. For example, the relation $\geqq$ on numbers is reflexive and the relation "in the same family as" is reflexive.

Another example of a partial order is implicit in the treatment of the comparative efficiency of mental tests on a "cost-utility" basis (1). "Cost" is the fraction of potentially successful people who are eliminated by a test; "utility" is the fraction of potential failures who are eliminated by the test. If for their respective cutting scores one test has a higher utility and a lower cost than another it is a superior test, but if it had a higher utility *and* a higher cost the two tests would be incomparable unless the relative weight of excluding a potential success to including a potential failure were known.

A basic problem in the theory of testing hypotheses in statistical inference is to test a simple hypothesis, H_0 (null hypothesis), against a single alternative hypothesis, H_1, by means of experimental data. A test, T, associates to each experimental outcome the decision to accept H_0 or to accept H_1 (but not both!). Each test T is appraised by a pair of numbers, namely, the probability of accepting H_1 if H_0 is true, $P_T(H_1|H_0)$, and the probability of accepting H_0 if H_1 is true, $P_T(H_0|H_1)$. Given two tests, T' and T'', then T' is said to be as good as T'' ($T' \geqq T''$) if and only if

$$P_{T'}(H_1|H_0) \leqq P_{T''}(H_1|H_0)$$
$$P_{T'}(H_0|H_1) \leqq P_{T''}(H_0|H_1)$$

The relation $\geqq$ on the set of all tests is an example of a partial order.

B_4, *lattice.* Let B be a poset relative to a relation $\geqq$. If a, b, c are elements of B and $c \geqq a$, $c \geqq b$, we say that c is an *upper bound* of a and b. If also $c \leqq x$ for every upper bound x of a and b, we say that c is the *least upper bound* of a and b and write $c = a \cup b$. In terms of the example of mental testing, c could be a person who passed exactly those items which were passed by at least one of a and b. Analogously, if $d \leqq a$, $d \leqq b$, we say that d is a *lower bound* of a and b, and if also $d \geqq y$ for all lower bounds y of a and b, we say that d is the *greatest lower bound* of a and b and write $d = a \cap b$. In our example d could be a person who passed exactly those items passed by both a and b.

A pair a, b need not have a least upper bound or a greatest lower bound. For example, if a passed items 1, 2, 3, 4; b passed 1, 2, 5, 6; c passed 1, 2, 3, 4, 5, 6, 7; c' passed 1, 2, 3, 4, 5, 6, 8; and there are no other persons, then both c and c' are upper bounds to a and b but there is no least upper bound. Also in this case there are no lower bounds for a and b and hence no greatest lower bound.

A poset is said to be a *lattice* if, for every pair a, b, both $a \cup b$ and $a \cap b$ exist. The lattice is an intermediate model between a partial order and a vector space.

George Miller[6] (Massachusetts Institute of Technology) has recently investigated the use of a lattice theoretic treatment of information in experimental psychology. To each item of information he associates (process A) an element of a lattice. If two items of information are associated respectively with elements x and y of the lattice, then the item which consists of the information common to the original items is associated with the element $x \cap y$, and the item which consists of the information contained in either of the original items is associated with the element $x \cup y$. As used by Miller, an item of information might consist of a cue or a sequence of cues in an experimental situation. The common procedure is to summarize the structure of the experiment by means of a lattice, given an experimental setup. However, the

[6] Personal communication.

abstract lattice in turn can motivate new types of experimental situations and indicate analogies between experimental designs which otherwise would not be apparent.

$B_{4'}$, *weak order.* A transitive order $\leqq$ is defined on $B_{4'}$ and has the property that for every pair a, b either $a \leqq b$ or $b \leqq a$. If both $a \leqq b$ and $b \leqq a$, we say that a and b are indifferent. Indifference is an equivalence relation (i.e., is reflexive, symmetric, and transitive).

A weak ordering would be illustrated by the military ranks of second lieutenant, first lieutenant, captain, major, etc. Each of these would constitute an equivalence class, and for any two officers (a, b), either $a \geqq b$ or $b \geqq a$, or both.

B_5, *chain.* A poset in which every pair is comparable is called a chain (or simple order, or linear order, or complete order). Alternatively, a chain is a weak order in which each indifference class consists of a single element. Here every pair of elements is ordered.

The previous example of a weak ordering of military rank could be converted into a chain if date of rank, standing in class, etc. were taken into account. Then, for every two distinct elements, a, b, either $a > b$ or $b > a$.

The ordinal scales of classical measurement theory are examples of chains.

$B_{5'}$, *partly ordered vector space.* A special case of lattice is provided by a real vector space (or a subset of a vector space). A *vector* $x = (x_1, \cdots, x_n)$ is an ordered set of n real numbers called the *components* of the vector. We define $x \leqq y$ to mean that $x_i \leqq y_i$ for each component. (Here the second symbol $\leqq$ refers to the usual ordering of real numbers.) This definition makes the vector space into a poset, and this poset is a lattice which is called a partly ordered vector space.

A partly ordered vector space is illustrated by the comparability of individuals in mental abilities. Conceiving of intelligence as made up of a number of primary mental abilities, each of these constitutes a component or dimension. Then, it may be said of two individuals x and y that y is at least as intelligent as x if and only if y has as much or more of each component as x has.

The term *vector* is sometimes used in a more general situation. If $C_1, \cdots, C_n$ are chain orders we may consider vectors or n-tuples $c = (c_1, \cdots, c_n)$ where the ith component c_i lies in the chain order C_i $(i = 1, \cdots, n)$. The set C of all such vectors c is called the *Cartesian product* of $C_1, \cdots, C_n$ and is denoted by $C = C_1 \times \cdots \times C_n$. We can make C into a poset by a process analogous to that used above for real vector spaces. Note that a real vector space is the special case of a Cartesian product of n factors $C_1, \cdots, C_n$ each equal to the set of real numbers.

B_6, *simply ordered vector space, or utility space.* A real vector space (or subset) in which $x < y$ is defined lexicographically, i.e., $x < y$ if $x_1 = y_1, \cdots, x_{i-1} = y_{i-1}, x_i < y_i$ is a special case of simple order.

A lexicographic ordering can be illustrated by the manner in which we might expect a fortune hunter to simply order a number of unmarried women. Presumably, financial assets would be the principal component and he might construct a weak ordering of the women, into, say five classes, on this basis. Then he would turn to the second component, say beauty, and within each of the financial classes construct a simple ordering of the women on this component. Any two women (a, b) would then be simply ordered as follows:

1. If a were in a higher financial class than b, a would be preferred to b.

2. If a were in the same financial class as b, then preference would be determined by their relation on the beauty component.

B_7, *real numbers.* The transition to scales using the real numbers has been given additional importance by the development of von Neumann-Morgenstern utilities. Even though one may wish to arrive here in order to have a simple index when a decision is to be made, it may frequently be desirable not to get here all at once, but to keep the components at a weaker level until it is necessary to map into the real numbers.

Measurement scales involving the real numbers, the interval scale, and the ratio scale have been discussed at length in the literature (**3, 5, 9, 10, 11**) and will not be pursued again here.

FURTHER EXTENSIONS

The various mathematical systems discussed here as available for measurement have been illustrated with objects of the real world mapped into the elements of an abstract system. A further level of abstraction is provided by defining a "distance function" in the abstract system, in which ordered pairs of elements in the abstract system are mapped into elements of another abstract system about which a variety of assertions may be made. In the context of measurement, these pairs of elements may correspond to "differences" between pairs of objects in the real world. These differences may themselves then be mapped into an appropriate abstract system such as one of those discussed here.

A number of these types of scales have been discussed by one of the authors (**4**, ch. 1). An illustration is the ordered metric scale in which the objects themselves satisfy B_5, a simply ordered scale, and ordered pairs of objects, regarded as "distances" between them, satisfy B_3, a partly ordered scale. Such scales are now being utilized for the measurement of utility and psychological probability in experiments on decision making under uncertainty (**6, 7**).

SUMMARY

One role of mathematical models is to provide a logical route to go from characteristics of the real world to predictions about it. The alternative route is by observation or experiment on the real world itself. The view expressed here is that these two routes are coordinate.

The various scales used in measurement serve as an illustration of the application of mathematical models and are subject to the same constraints as other mathematical models. That is, if the axioms underlying the scale are not satisfied by that segment of the real world which is mapped into it, then the interpretations of the mathematical conclusions may have no reality or meaning. Thus, to insist that measurement always constitutes the mapping of physical objects into the real number system is to impose on the real world an abstract theory which may be invalid.

A partial ordering of various alternative mathematical systems available for measurement has been presented with illustrations in order to reveal the relative strengths of these scales to which the real world must conform to permit their application. We make no claim to completeness in this list of models for measurement theory. Our purpose is to point out the richness of the set of possible models and to give some examples that show how the use of more general models can extend the domain of classical measurement theory.

None of the discussion here should be taken as an argument for the use of weaker scales in the place of stronger scales for their own sake. The measurement scale utilized constitutes a theory about the real world and the stronger the theory the better, so long as it is correct. The addition to a scale of axioms which are not satisfied by the real world is a step away from the path of progress.

REFERENCES

1. BERKSON, J. "Cost-utility" as a measure of the efficiency of a test. *J. Amer. stat. Ass.*, 1947, 42, 246–255.
2. BIRKHOFF, G. *Lattice theory.* New York: American Mathematical Society, 1948.
3. COOMBS, C. H. Mathematical models in psychological scaling. *J. Amer. stat. Ass.*, 1951, 46, 480–489.
4. COOMBS, C. H. A theory of psychological scaling. *Engng Res. Bull.* (Univer. of Michigan), 1952, No. 34.
5. COOMBS, C. H. The theory and methods of social measurement. In L. Festinger & D. Katz (Eds.), *Research methods in the behavioral sciences.* New York: Dryden Press, 1953. Pp. 471–535.
6. COOMBS, C. H. Social choice and strength of preference. In C. H. Coombs, R. M. Thrall, & R. L. Davis (Eds.), *Decision processes.* New York: Wiley, in press.
7. COOMBS, C. H., & BEARDSLEE, D. C. On decision making under uncertainty. In C. H. Coombs, R. M. Thrall, & R. L. Davis (Eds.), *Decision processes.* New York: Wiley, in press.
8. KERSHNER, R. B., & WILCOX, L. R. *The anatomy of mathematics.* New York: Ronald, 1950.
9. STEVENS, S. S. On the theory of scales of measurement. *Science*, 1946, 103, 677–680.
10. STEVENS, S. S. Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1950. Pp. 1–49.
11. WEITZENHOFFER, A. M. Mathematical structures and psychological measurements. *Psychometrika*, 1951, 16, 387–406.
12. WEYL, H. *Philosophy of mathematics and natural science.* Princeton: Princeton Univer. Press, 1949.
13. WILDER, R. L. *Introduction to the foundations of mathematics.* New York: Wiley, 1952.

(Received June 1, 1953)

VOL. 61, NO. 4 JULY, 1954

THE PSYCHOLOGICAL REVIEW

THE RELATION OF RESPONSE LATENCY AND SPEED TO THE INTERVENING VARIABLES AND *N* IN S–R THEORY

KENNETH W. SPENCE

State University of Iowa

No 13. May [illegible]

I

In the *Principles of Behavior* Hull introduced his theoretical constructs (intervening variables) initially in terms of independent environmental variables (e.g., S_e, N, T_d, etc.), and completed the theoretical structure by anchoring them to certain response measures. The latter involved the introduction of specific *ad hoc* postulates that related each of the response measures (e.g., latency, frequency, resistance to extinction, etc.) to one or other of the theoretical constructs. Thus in the case of the response measure (latency) with which we shall be concerned, Hull made the following assumption:

> The latency of response (R_t) is a decreasing hyperbolic function of the momentary effective excitatory potential ($\dot{\bar{E}}$), i.e., $R_t = a\dot{\bar{E}}^{-b}$, where a and b are empirical parameters (Postulate 13, p. 344).

Attention has been called elsewhere (7, 8) to the point that certain of these postulates are entirely superfluous in that assumptions already a part of the system (i.e., earlier postulates) permit one to derive a necessary relation between these response measures and one or other of the theoretical constructs. Thus in the case of the postulate introducing response probability or frequency ($R\%$) as a function of effective excitatory potential ($\bar{E}$), the relation assumed is, as has been shown, derivable from definitions and postulates already made concerning effective excitatory potential ($\bar{E}$), oscillatory inhibition (O) and reaction threshold (L). In this particular instance, the postulate Hull made (normal integral function) happened to be identical with the relation that may be derived from assumptions already a part of the system (6). In the case of the latency measure, however, it may be shown that the postulate he assumed is actually inconsistent with a necessary relation that follows from earlier assumptions. The main purpose of the present article is to consider the implications for this relationship of the assumptions already made concerning $\bar{E}$, O, and L, and to extend their implications to the empirical laws (learning curves) to be expected between response latency and speed, on the one hand, and the variable N on the other.[1]

We need to recall first that momentary effective excitatory potential, $\dot{\bar{E}}$, is equal to $\bar{E} - O$, and that a response

[1] The assumptions made concerning O are those given in Hull's *Principles of Behavior* (3), not those in his later *Essentials of Behavior* (4) and *A Behavior System* (5).

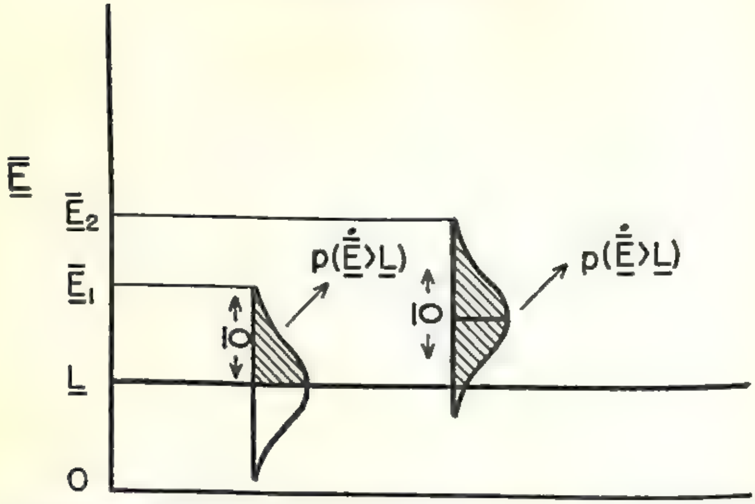

FIG. 1. The shaded portions of the upended normal distribution functions show the probability of $\dot{\bar{E}}$ being greater than L for two levels of $\bar{E}$

is assumed to occur to a stimulus only when (*a*) $\bar{E}$ is greater than L, and (*b*) when an O value exists that is sufficiently small to make the value of $\dot{\bar{E}}$ greater than L. Oscillatory inhibition (O), it will be remembered, is assumed to change in value from moment to moment, the distribution of the values being postulated as normally distributed. The problem becomes one, then, of determining the average time, $\bar{t}$, before a momentary O value occurs that will provide an $\dot{\bar{E}}$ value greater than L.

Let P be the probability of occurrence of such an O value. Then the probability that such a value of O will be the first one to occur is P; the probability that such a value will be the second is $(1-P)P$; the probability that it will be the third is $(1-P)(1-P)P$ or $P(1-P)^2$, etc. Considering now an indefinitely large number of occasions on which the stimulus is presented, and representing the number of occurrences of momentary O values on any occasion by n, we may weight each possible value of n by its probability (expected relative frequency), and thus obtain a mean expected value of n. Estes (1) has shown that this mean value, $\bar{n}$, is equal to $1/P$. In other words, $\bar{n}$ is the mean expected number of momentary O values that will occur on each stimulus occasion until an O value that provides a superthreshold $\dot{\bar{E}}$ value will occur.

If now we let u' represent the average time or duration of a momentary O, then the average time, $\bar{t}$, for a superthreshold $\dot{\bar{E}}$ value to occur will be the product of the expected number of momentary O values that will occur and the mean duration of a momentary O.

$$\bar{t} = \bar{n}u' = \frac{u'}{P}. \qquad (1)$$

In terms of an average measure of speed of response evocation ($\bar{v}$), this equation becomes

$$\bar{v} = 1/\bar{t} = \frac{P}{u'} = uP, \qquad (2)$$

where

$$u = 1/u'.$$

Figure 1 represents two levels of $\bar{E}$ and shows their relation to L and O. The probability (P) of O being a value that will produce a superthreshold $\dot{\bar{E}}$, which can also be described as the probability that $\dot{\bar{E}}$ is superthreshold, $[p(\dot{\bar{E}} > L)]$, is given by the proportion of the upended normal distribution that is above L. This yields

$$P = p(\dot{\bar{E}} > L) = \int_{-\infty}^{\bar{E}-L-2.5\sigma O} (O)\, dO. \quad (3)$$

Since $\bar{E}$ is assumed to be an exponential function of N, i.e., $\bar{E} = A(1 - e^{-iN})$, it is possible to ascertain the theoretical function that P is of N by means of a table of cumulative probability values of the normal curve. Figure 2 shows the family of theoretical curves of the proportion of superthreshold $\dot{\bar{E}}$ values, $p(\dot{\bar{E}} > L)$ as a function of N for different curves of growth of $\bar{E}$. When multiplied by the parameter, u, representing the reciprocal of average duration of a momentary O value, such curves provide a theoretical prediction of $\bar{v}$ as a function of N. The relationship is identical in form, it should be noted, with the theoretical frequency measure in classical conditioning (7) and implies an initial period of positive acceleration, providing that the data represent the major portion of the total possible learning and not just some later part of it.

Now the measure $\bar{l}$, it should be noted, is a measure of the time it takes to get the effector activity initiated. As such it does not involve the time or duration of whatever neuromuscular activity is involved on the part of the subject in the measuring situation. In actual practice, however, any measurement of response latency must also involve the time taken by the action of the effector system. Presumably our so-called latency measures, or measures of response time, represent a summation of these two durations so that, if we let T represent the obtained experimental measure of response time, $\bar{l}$ the measure of action latency, and t' the duration of the effector activity involved in the meas-

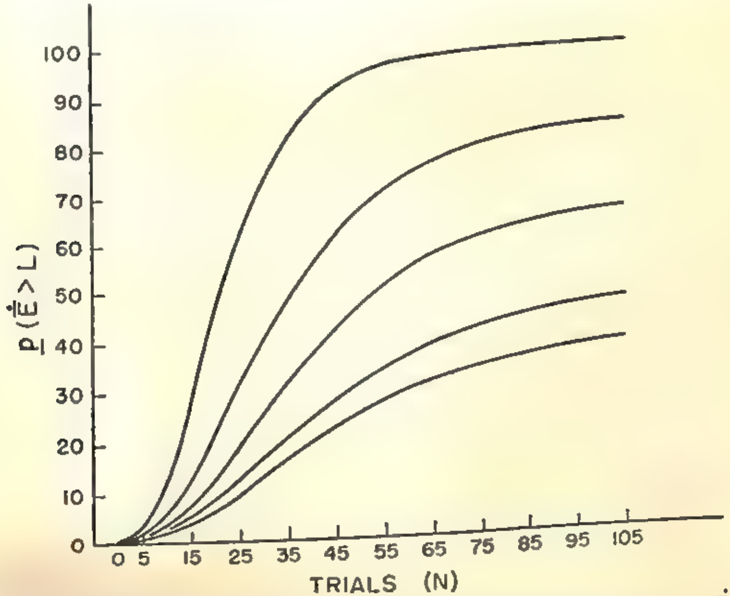

FIG. 2. Family of theoretical curves of the proportion of superthreshold $\dot{\bar{E}}$ values as a function of N for different curves of growth of $\bar{E}$

uring operation, then

$$T = \bar{t} + t'. \tag{4}$$

A similar equation can be derived for a measure of speed of response (V) as follows:

$$V = \frac{1}{\bar{t} + t'}. \tag{5}$$

The problem immediately arises as to how t' (and its reciprocal, v') vary with $\bar{E}$. The present writer has not been able to derive a relation between t' and $\bar{E}$ from any of the existing postulates of the system. Accordingly, the working hypothesis is made that the relation is the simple hyperbolic one shown in the following equation:[2]

$$t' = \frac{c}{(\bar{E} - L)} \quad \text{where } c \text{ is a constant.}$$

Substituting now in equations 4 and 5, we obtain the following equations as representing the functions relating the experimental measurements of time (T) and speed of response (V) in simple instrumental conditioning to $\bar{E}$:

$$T = \frac{u'}{P} + \frac{c}{(\bar{E} - L)}, \tag{6}$$

$$V = \frac{1}{\dfrac{u'}{P} + \dfrac{c}{(\bar{E} - L)}}. \tag{7}$$

An interesting implication of the above theorizing is that the shape of the curve of V as a function of $\bar{E}$, and hence of N, will depend upon the magnitude of the parameter c, which is experimentally manipulable by varying the amount (duration) of motor activity involved in the measurement of the response. Thus in the simple approach type of situation (locomotion in a straight alley), one could vary the value of c by measuring the response for different lengths of runway. When a minimum length of alley was employed, c would approach zero, and equation 7 would become identical with equation 2. Under this condition, the speed measure would provide an initially positively accelerated curve.

As the length of the runway involved in the measurement is increased, the value of c will increase, and the family of speed-of-response (v) curves as a function of N would be expected to vary in their initial phase from positive to negative acceleration. By selecting an appropriate length of runway, one should be able to obtain a curve that is linear in its early course of development. Finally, it should be noted that the theory implies that if the measurement (t') is taken from a point after the activity has started, the curve for speed of running (v') as a function of N should be a negatively accelerated exponential function, i.e.,

$$v' = \frac{A(1 - e^{-iN}) - L}{c}. \tag{8}$$

II

The above derivations concerning the relation of the measures, response time and speed of response to N in instrumental learning, assume that the growth of $\bar{E}$ is an exponential one of the type that Hull employed. This assumption, in turn, depends upon the postulate that H grows in this manner and that the other factors determining $\bar{E}$, such as stimulus dynamism Q, drive D, incentive motivation K, and work inhibition I, are *constant* throughout the course of the training period. By means of various experimental techniques, such as distributing the trials, having only a few trials a day (3 or 4), etc., it is possible to keep Q, D, and I fairly constant. The situation is not so simple so far as the incentive moti-

[2] This implies, of course, that the relation between speed of activity (v') and $\bar{E}$ is linear. $v' = \frac{(\bar{E} - L)}{c}$.

vational factor, K, is concerned. In recent discussions of theories of learning (6) the writer has suggested that this factor K might represent a stimulus dynamism, that provided by the proprioceptive component (s_G) of the fractional anticipatory goal response (r_G). According to this notion, in instrumental learning involving reward, one also has classical conditioning taking place so that a fractional part of the goal response becomes conditioned to the stimulus situation. With conditioning, this r_G and its cue s_G, by virtue of generalization, move forward in time and thus become a part of the internal stimulus complex determining the strength of the instrumental response. Thus it is a kind of acquired motivating factor, the strength of which depends not only on the conditions of reinforcement (i.e., magnitude and delay of reinforcement), but also on the stage of training, i.e., number of reinforced trials (N).

This hypothesis has a number of important implications, not only for experiments involving different magnitudes and delays of reinforcement, but also for the nature of the learning (performance) curve in simple instrumental learning. Confining our interest here to the latter problem, we shall consider the implication of the variation of K during learning *without taking into account how this motivational variable interacts or combines with the other two motivational constructs, Q and D.*

According to our assumption concerning E, and ignoring D and Q,

$$R = f(E) = (K \times H). \qquad (9)$$

Substituting for K and H their postulated relations to N, we obtain the following:

$$R = f(E) = [B - (B - x)e^{-aN}] \times [A - (A - x)e^{-iN}]. \qquad (10)$$

According to equation 9, the growth of E as a function of N would be initially positively accelerated instead of the negatively accelerated exponential function that would obtain if K were some constant value. In view of the fact that the curves of learning, e.g., $V = f(N)$ and $v' = f(N)$, are determined in part by the growth of E, it is readily apparent that we need to take this function into account in making any predictions as to the form of these curves. Thus, one interesting implication of this hypothesis is that if one were to establish K at a maximum (hence constant) value by setting up the classical CR ($S_c - r_G$) to the sight and sound of the lever (S_c) *prior to the beginning of the instrumental learning*, the growth of E would now be expected to be a negatively accelerated function throughout its course. Under this condition, the period of positive acceleration of the speed of response evocation curve $\bar{v} = f(N)$ should, other things being equated, be less than under the normal procedure in which the classical conditioning proceeds along with the instrumental learning.

In a similar fashion, one may predict that curves of speed of running ($v' = fN$) would be a negative growth function only under the condition in which K is a constant, and would tend to exhibit an initial phase of positive acceleration to the extent that K grows from zero to its maximum throughout the course of learning the instrumental response. It is probably the case that K is, as the result of transfer from past experience, already considerably developed in the simple running situation under normal illumination conditions, with the consequence that the curve of growth of E is distorted only slightly from the negative exponential function that holds when K is constant.

III

In this section some of the problems connected with the experimental testing of these theoretical implications will be discussed. If one examines the existing data from instrumental learning investigations, it will be found that a variety of curves of speed of response (reciprocal of latency or response duration) have been obtained. Of some score of curves, both from the literature and from unpublished studies conducted in the Iowa laboratory examined by the writer, it was evident that the majority showed a relatively brief initial period of positive acceleration followed by a period of prolonged negative acceleration to the asymptote. The second most frequently observed type was linear in its early portion (10 to 15 trials), followed by a negatively accelerated approach to the limit. A few curves exhibited negative acceleration throughout their course. It is the writer's impression that the latter type of curve tends to occur when the drive is very high, such as in running to escape an electric shock or under a strong hunger drive.

Unfortunately, these data are not very satisfactory for the reason that they are typically group curves involving the mean or median of a whole group of subjects. The forms taken by such group curves may deviate markedly from the curves of individual subjects, as Hayes (2) has recently so nicely demonstrated. The reasons for this become obvious when we consider that there are marked differences among individual subjects, not only in initial and final levels of performance, but also in their different rates of learning, i.e., their different rates of approach to the performance asymptote.

A number of alternatives to the use of such group curves suggest themselves. One is to employ the data of individual subjects. The notorious variability of individual measures from trial to trial, however, usually necessitates some form of averaging of the individual measures in terms of blocks of trials, a procedure which also often leads to distortion of the curve. Thus if an individual curve that shows an initial phase of positive acceleration within the first 10 trials followed by a negatively accelerated phase is averaged in terms of successive blocks of 10 trials each, the initial period of positive acceleration will be lost entirely. The most satisfactory manner of treating individual measures, particularly measures of speed of response, would appear to be the moving average method with small blocks of trials, e.g., three trials in a block.

As an alternative to these curves based on individual measures and averages of a group of subjects, the writer has employed curves based on "like," or homogeneous, subjects. The homogeneity of the subjects can be ascertained in terms of the likeness of the subjects' performances at different stages of the practice or throughout the total learning period. Thus, in the case of the frequency measure in classical conditioning, the subjects' scores in terms of the total number of CR's occurring in a given number of trials, e.g., 100, can first be determined, and then groups of "like" subjects can be formed in terms of those that fall in a small range of scores, such as from 20 to 30 CR's, or from 50 to 60, etc. The writer has shown that such data from different parts of the distribution of total CR scores provide very smooth, comparable curves with relatively small numbers of subjects. In such curves the form is not a function of the distribution of the individual scores.

There are, of course, further refinements that can be made in this procedure. For example, in terms of the speed measures we have been discussing, one could obtain measures for each subject at the beginning, at some intermediate point, and at the end of the learning, and then form groups of subjects that are alike at all three points rather than alike only on the basis of an over-all performance measure. As psychology moves into a period when testing of its theories requires the evaluation of the empirical data in terms of the precise form of some predicted lawful relation, more refined procedures for ascertaining the nature of the function will have to be made available.

A second point in connection with the testing of these theoretical implications is that the experimenter must define his response measures more precisely. Thus the present theory makes it necessary to differentiate between measures that involve starting of the action, the duration of the activity once set going, and combinations of both. The predicted differences in these various measures provide one of the most feasible ways of testing the theory. Similarly, the motivational conditions, relevant and irrelevant, will have to be carefully controlled and their possible differential effects taken into consideration.

Undoubtedly, the most difficult task will be that of arranging the experimental conditions so that there are no competing responses in the situation; for the theoretical model assumes but a single response, not a number of competing responses. Most instances of instrumental conditioning are really limiting cases of trial-and-error learning in which competing responses, while minimized, still play a more or less important role. Obviously, the occurrence of other competing responses in such simple learning situations involves time and thus importantly affects speed measures. Moreover, it is an unfortunate fact that interference from such competing responses is greatest at the beginning of instrumental learning when the rewarded response is weak. Later in the learning, this response is so much stronger than the competing ones that the latter are unable to interfere.

Elimination of competing responses will necessitate both experimental procedures and objective criteria for eliminating data on trials on which competing responses do occur despite the experimental controls. Particularly important also is control of the response orientation of the subject just prior to the presentation of the stimulus. A procedure that has been found to be quite successful in obtaining such control in a very few trials (two to four) is to employ two doors, one opaque and the other glass, in the starting box of an operant situation such as the straight alley. Each trial is started by the experimenter raising the opaque door first and then, at a fixed interval (three seconds) thereafter, the glass door. The raising of the opaque door serves as a warning signal for the subject, which very quickly comes to be set to respond to the raising of the glass door. Similarly, measuring techniques will have to be as precise as possible, particularly in measuring the speed of response evocation. The stop watch will hardly serve for our present purposes.

Finally, in addition to controlling competing responses, it will be necessary to minimize transfer from past experience so that the strength of the S-R is sufficiently low to provide a picture of the major portion of the total learning. It is particularly important from the point of view of the present theory that a good share of the initial phase of learning be represented in the data.

IV

In the preceding sections we have elaborated some of the implications with respect to the form of the speed of response evocation curve of learning of the original theoretical model put forward by Hull in his *Principles of Behavior*. The present treatment differs somewhat from that given by Hull in that certain implications of his postulates with respect to excitatory potential ($\bar{E}$), oscillatory inhibition (O), and threshold (L) were developed, with the consequence that it was possible to derive the relation to be expected between speed of response evocation ($\bar{v}$) and $\bar{E}$. By means of an assumption relating speed of running (v') to $\bar{E}$, additional implications were drawn concerning measures that involved various combinations of the two measures, speed of response evocation ($\bar{v}$) and speed of running (v'). Also considered were certain problems relating to the obtainment and treatment of experimental data bearing on the theory.

REFERENCES

1. Estes, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, 57, 94–107.
2. Hayes, K. J. The backward curve: a method for the study of learning. *Psychol. Rev.*, 1953, 60, 269–276.
3. Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
4. Hull, C. L. *Essentials of behavior.* New Haven: Yale Univer. Press, 1951.
5. Hull, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.
6. Spence, K. W. Theories of learning. In C. P. Stone (Ed.), *Comparative psychology.* (3rd Ed.) New York: Prentice-Hall, 1951.
7. Spence, K. W. Mathematical formulations of learning phenomena. *J. exp. Psychol.*, 1952, 59, 152–160.
8. Spence, K. W. *Symposium on relationships among learning theory, personality theory, and clinical research.* New York: Wiley, 1953.

(Received November 25, 1953)

Psychological Review
Vol. 61, No. 4, 1954

CONDITIONING AS AN ARTIFACT

KENDON SMITH

The Pennsylvania State University[1]

The case for a pure reinforcement theory of learning has been strongly put by recent papers (20, 21, 26, 40), and it is difficult now to escape the conviction that such a view is essentially correct.

In spite of its fundamental strength, however, reinforcement theory has remained weak in one respect. It has experienced continual difficulty in handling the problem of autonomic, visceral learning. To be consistent, Hull was obliged to maintain that reinforcement was crucial even in the presumed acquisition of such responses as salivation, alteration of skin resistance, cardiovascular changes, etc. This he did (19, pp. 76–80), although not as flatly as he is sometimes said to have done. Hull thus arrived at a position that many have regarded as untenable; for, although it is possible to imagine reinforcemental factors at work in some instances of alleged visceral learning, there are many other instances in which the influence of such factors seems to be completely out of the question.

One variant of reinforcement theory that is designed to meet the situation more adequately is the so-called "two-factor" theory. It has been most recently and most vigorously promoted by Mowrer (29, 30), but Mowrer's views are in reality quite similar to those previously advanced by Thorndike (44, pp. 401–412) and Skinner (39, pp. 109–115). Mowrer differentiates between *conditioning* and *problem solving*. The first term refers to the process of learning by a stimulus-substitution, contiguity principle; the second term, to that of learning by a principle of reinforcement. Conditioning is alleged to occur specifically via the autonomic nervous system, and problem solving to be mediated solely by the "central" (i.e., somatic) nervous system; a sharp distinction between the two neurological divisions is drawn and emphasized.

Such incisive dualism has an undeniable appeal. It suffers, however, from a disquieting lack of parsimony. It is extremely difficult to believe that "visceral learning" and somatic learning, so alike in so many respects, must be accomplished by two different principles (cf. 38). The resolution of the problem thus appears to lie elsewhere.

As it happens, there exists a rather simple way to save reinforcement, and parsimony with it. One can begin by accepting the notion that somatic learning is reinforcemental learning. He can then go one step further than Mowrer has, and expunge from the viscera not only problem solving but conditioning as well. This procedure leaves the law of effect to rule in monistic majesty. Of course, it also makes the autonomic nervous system totally uneducable; but that, it can be asserted, is as it should be. For it can be argued that every "conditioned visceral response" is in reality an artifact, an innate accompaniment of the skeletal responses inculcated by the conditioning process.

The present paper will, in fact, attempt to defend this general line of thought.

[1] The writer wishes to acknowledge with real appreciation the helpful and sustained interest of his colleagues, Professors Joseph H. Grosslight, John F. Hall, and William M. Lepley.

THE ARGUMENT AND A SPECIFIC INSTANCE

In the elaboration of this proposal, it might be wise to begin with a concrete example. The galvanic skin response and its "conditioning" would seem to be appropriate.

When an attempt is made to condition the galvanic skin response, the procedure generally pairs a neutral conditioned stimulus with an unconditioned stimulus that is more or less noxious in nature, typically an electrical shock. Several pairings will eventuate, in some subjects, in a new correlation between the GSR and the originally neutral stimulus. "Conditioning"[2] has occurred.

Anyone who has actually carried out this procedure knows, however, that the foregoing account is seriously incomplete. The subject is not by any means a passive hulk during the entire program. In particular, he soon comes to regard the conditioned stimulus as a signal for a muscular "bracing" against the noxious stimulus to come; presumably this bracing, being somatic, is a matter of reinforcemental learning. At any rate, one would expect such skeletal activity to be accompanied by the GSR, simply as a matter of innate neural connections. The occurrence of the conditioned GSR is thus hardly surprising, and it can be explained without recourse to a principle of autonomic learning. It is, in short, an artifact.

Now, at least two objections arise immediately. The first is that a scheme that works so well for one response, the GSR, may not work at all for others: what about salivation, for instance? This is a reasonable misgiving, but it will be reserved for later consideration. The second objection is basic and serious, and it will be faced immediately.

The difficulty is this: Although it may seem quite natural to think of somatic responses as somehow "causing" their concomitant autonomic reactions, and although such a notion frequently finds expression in the literature, it is well to remember that the causal sequence is actually most obscure. It is entirely possible, for instance, that the muscular tension and its associated GSR arise as parallel events from a common innervation. The conservative view would seem to be that there is no distinguishable priority as between the two responses. By what right, then, is one considered the "real" response and the other a mere by-product?

Common sense would very likely answer that question rather readily: "The muscular response is the real response, because it is *conscious;* I didn't even know there was a 'galvanic skin response' until you told me about it!" Common sense thus suggests an answer which makes a certain amount of scientific sense too.

The combined, bracing-GSR reaction has already been pictured as a unitary one; and it was agreed to begin with that somatic responses attach themselves to new stimuli according to a reinforcemental paradigm. It follows, therefore, that the combined response, as a unit, gets about by reinforcement. Now, the acquisition and maintenance of a voluntary (i.e., reinforcementally acquired) response is dependent upon the existence of afferent cues (cf. 4, pp. 524 ff.).[3] The skeletal re-

[2] Inasmuch as this paper ultimately arrives at the conclusion that "conditioning" does not actually exist, the term itself has been treated so far in a rather gingerly fashion. In the interests of clean typography, quotation marks will be omitted from now on; but when conditioning and similar terms are used, they are meant to be read as if quotation marks were still present.

[3] It is perhaps possible that this dependence stands as a testimony to the importance of immediate higher-order reinforcement; the af-

sponses, in the case at hand, provide a wealth of afferent information; but the autonomic reactions generate no regulatory feedback whatsoever. Acquisition of the whole response pattern, therefore, would seem to depend upon the integrity of the somatic component. If it were not for the muscular response, the GSR would not exist; but the GSR could be eliminated, perhaps by sympathectomy, and the bracing response would remain unaffected. The GSR is truly, in this sense, a secondary phenomenon, a by-product.

The earlier analysis of the conditioned GSR thus appears to be valid. At the same time, examination of this particular instance has generated a logic that can be applied to other instances of alleged autonomic learning. If it can be shown that the development of any new autonomic response is coincident with the growth of a somatic response known to have such an autonomic response as a regular correlate, it is legitimate to label the autonomic response as a secondary effect and the conditioning as bogus.[4]

The question remaining, then, is this: Do other instances of conditioning fit this pattern well enough to permit generalization of the artifact hypothesis?

APPLICATIONS IN OTHER INSTANCES

As one might expect, some varieties of conditioning conform to the pattern quite obviously, and others do not. For example, there are quite a few autonomic responses which, like the GSR, are known to be associated with the diffuse skeletal reactions that develop during a conditioning procedure. The foregoing discussion might just as well have revolved about the response of vasoconstriction as about the GSR; thus, conditioned vasoconstriction (2, 25, 37) fits the theory well. The same might be said for the few reported instances of conditioned vasodilation (2, 25), and for conditioned cardiac deceleration, which was recently observed in animals and human beings (22, 31). On the other hand, there are also instances that do not fall into place quite so neatly. Nevertheless (and this is the burden of the paragraphs to come), they do not, on close examination, present the prohibitive difficulties one might expect.

A case in point is that which generated the concept of conditioning in the first place: the salivary reflex. Everyone knows about Pavlov's classical experiments, and everyone knows that Pavlov induced his experimental animals to salivate on signal. It is not quite so widely known, however, that the animals did several other things besides salivating. Of special interest at the moment is the fact that they displayed gross skeletal responses. Pavlov observed movements of the head and "smacking . . . [of the] lips" (32, pp. 29–30), the animal appearing ". . . to take the air into its mouth, or to eat the sound" of the conditioned stimulus (33). Pavlov spoke frequently of the "alimentary reflex" or the "complex reflex of nutrition" (32, cf. pp. 13–14), and emphasized the fact that he was dealing with a pervasive pattern of behavior rather than with salivation alone.

These facts are manifestly made to order for the present hypothesis. Further, they have been confirmed by later experimenters. Zener, who also worked with the salivary reflex in dogs (48),

ferent cues may constitute particularly prompt rewards and punishments. In any event, the dependence seems to be a fact.

[4] It is true that some visceral responses arouse afferent neural activity. As a class, however, these responses are sluggish, and the afferent information that they provide is greatly delayed in returning to the central nervous system. It can be presumed that there might as well be no return at all.

sometimes saw "chewing and licking" responses when the conditioned stimulus was administered; Zener and McCurdy (49) reported a correlation between rate of salivation and rate of chewing. And Moore and Marcuse, who stoutly undertook to condition the salivary responses of two sows (27), were explicitly concerned that the observed salivation might be due to oral activity. The records obtained by Moore and Marcuse seemed to give negative indications. These experimenters, nevertheless, had finally recognized the possibility, which had been so curiously neglected until their time, that the salivary responses were not essentially independent of the motor responses in the animal's behavior, but rather that the visceral activity of salivation was a natural concomitant of the acquired somatic activities of chewing, swallowing, etc. In this case, it might be noted, the possibility of concomitance had existed not only in the sense in which it has been developed in the foregoing discussion. It had existed also in the more obvious sense that sheer mechanical stimulation arising from oral activity could be expected to elicit salivation directly.

Of special interest at this point are Razran's reports of conditioned salivary responses in human subjects. Razran's experiments are well known, but it might be emphasized that, here once again, it was a complex process that was under investigation. Early results were rather widely irregular (34), and Razran soon came to recognize the importance of his subjects' "attitudes" (34, 35, 36). It turned out also that the act of thinking was important in the experimental results: ". . . it is seemingly not mere 'willing' but thinking of some more or less specific stimulus or response that produces a voluntary flow of saliva" (34, p. 9). Explicit instructions to "form associations" between conditioned and unconditioned stimuli "produced very effective positive conditioning," while instructions to avoid forming associations usually had an opposite effect. It was even possible to substitute the thought of eating for actual eating, as the *unconditioned* stimulus, and still have a successful experiment.

Razran cautioned his subjects against gross oral movements, and his injunctions were evidently obeyed (34); thus the skeletal responses that might have evoked salivation were not as obvious in this instance as they were in the case of animal conditioning. Nevertheless, the fact that thought processes seemed so important is quite suggestive. If there is anything at all to a motor theory of thought and imagination, one must acknowledge the possibility of oral activity similar in nature, if not in magnitude, to that of the animal subjects. If such activity occurred, it might well have evoked the measured salivation. To be emphasized, also, is the role of gross bodily tension. Razran's subjects were hungry, and the food signal might well have led to a certain degree of general muscular relaxation in anticipation of ingestion. Such relaxation could be expected to tip the autonomic balance toward parasympathetic dominance and thus toward salivation.

It appears, then, that the conditioned salivary response is not beyond reconciliation with the notions proposed earlier. Attention may now be turned to another standard example of visceral learning: the conditioned pupillary response, which is of special interest in the present context principally because it evidently does not exist.

THE PUPILLARY RESPONSE

It is, perhaps, surprising to discover that pupillary conditioning is even in doubt. The early reports of Cason (5) and Hudgins (17) were optimistic and well publicized. They mentioned not only iridic conditioning but even "voluntary control" (17) of the pupillary response. In 1934, Steckle and Renshaw reported an attempt to condition the pupillary reflex (42); the attempt was unsuccessful, and the experimenters expressed some skepticism about the earlier accounts. The ensuing discussion in the literature (18, 41) made it clear that these earlier studies had been replete with technical difficulties that left considerable room for subjective, judgmental factors. In 1936, in the second of the two papers last mentioned, Steckle again reported negative findings.

In 1938, positive results were claimed once more. Baker described iridic conditioning to "subliminal" and supraliminal sounds (1). Conditioning was alleged to be particularly rapid and stable when subliminal stimuli were employed. An elaborate repetition of Baker's work by Wedell, Taylor, and Skolnick (45) failed to confirm his results, as did a careful check by Hilgard, Miller, and Ohlson (16). A recent and thorough exploration by Hilgard, Dutton, and Helmick (14) has again produced very little evidence of successful iridic conditioning. Citing similar results with animals as well as with human beings (43), these latter authors warn that continued negative findings may force a revision of accepted learning theory.

In terms of the matter at hand, the defection of the pupillary response has a double significance. In the first place, it means that no theory need be seriously concerned with accounting for the existence of a conditioned iridic reflex; the present formulation, along with others, thus escapes a certain amount of travail. In the second place, and more importantly, the failure of the pupillary response is materially embarrassing to a conditioning theory of visceral learning. Here, in the hands of competent workers, the principle of contiguity has had ample opportunity to exhibit itself. It has not done so. If "contiguity" and "artifact" are regarded as exhaustive alternatives, the artifact hypothesis appears to be the sole survivor.

It is of some incidental interest to speculate as to why conditioning should fail in the special instance of the pupillary response. It might be suggested that the changes in level of illumination that are customarily employed in iridic-conditioning experiments are of no practical consequence to the subjects, and that, therefore, no anticipatory skeletal responses are acquired. There being no acquired skeletal responses, there is no conditioning.

In this connection, it is worth noting that unexplained failures characterize almost every conditioning experiment. Some subjects condition, and others do not. Such vagaries are much more suggestive of individual differences in personality structure than they are of the presumed essential similarity of fundamental neurological processes from person to person. One thinks of greater and lesser tendencies toward anticipation, anxiety, and (literal) tension as perhaps underlying the personal variations in conditionability. It is evident that the recently reported correlations between anxiety level and conditionability fit in rather neatly with such a conceptualization (3, 31, 46).

Tests and Testability

The foregoing evidence is taken to be strongly favorable to an artifact theory of conditioning. It hardly suffices to close the matter, however. It seems appropriate now to examine other lines of evidence, formal and informal, actual and potential, which might have a bearing upon the hypothesis at hand.

From everyday experience comes the first datum: it is notoriously possible to forestall an untoward visceral response by "not thinking about it." Stimuli that might otherwise elicit nausea, flushing, or sexual reflexes lose much of their effectiveness when one refuses to dwell on their implications. If thought is somatic, it would thus appear that the autonomic response depends upon the somatic and does not arise independently of it. There are clear and perhaps important implications here for a further understanding of such phenomena as hysterical nausea and psychological impotence.

A somewhat similar argument also arises from everyday experience. Sheffield (38) has well pointed out the obvious fact that bowel and bladder functions are in some sense subject to reinforcemental training. An equally obvious fact, to be emphasized here, is that such training is conducted in terms of massive somatic responses, and that everyday visceral control is exercised only by virtue of diffuse contractions of the skeletal musculature. Again, it appears that the "real" response is a somatic one.

Both of the preceding arguments bolster the basic contention that learned visceral activity is a by-product of somatic behavior. A crucial test of this contention could be made if one could provide oneself with a subject who is completely passive, and even unthinking. A subject in such a state should not, according to theory, be susceptible to conditioning. Could such a situation be contrived under deep hypnosis? One can imagine a good hypnotic subject, instructed to relax and to "keep his mind completely blank," but to remain awake and aware of such stimuli as bells and shocks; would the conditioned GSR appear in such circumstances?

To the best of the writer's knowledge, such an experiment has not been performed. It must be admitted, however, that a rather similar situation has been reported, and that it does not seem to conform to theoretical expectations. In 1938, Lindsley and Sassaman (24) reported the discovery of an individual who was able to exercise "voluntary control" over his pilomotor responses. The body hair could be erected or lowered upon signal. The subject reported having come upon his talent more or less accidentally. He denied that the basis for his performance was controlled imagery: he imagined nothing in particular; he simply "willed" the response. Neither were there obvious skeletal movements or muscular tensions to account for these hirsute accomplishments. This case, anomalous as it is, presents difficulties for practically any theory of conditioning. The only saving feature, as far as the present hypothesis is concerned, is that there were indications of a very diffuse activity of some sort: cardiac and pupillary effects accompanied the pilomotor responses. The possibility of a covert somatic response was thus considerable.

The tenor of this discussion suggests a well-known animal investigation, that of Harlow and Stagner (12). Using cats as subjects, these investigators paralyzed the striate musculature by deep curarization. While the animals were under curare, the pupillary response to electrical shock (which could still be elicited) was conditioned to the sound of a buzzer. Here again is ostensibly negative evidence. As it has turned out, however, the effects of curare upon the neuromuscular system seem to be quite complex, and later experiments have found striate-muscle responses even in deeply curarized animals (10, 11). If Harlow and Stagner's animals retained some degree of skeletal responsiveness, the experiment loses much of its crucial aspect as far as the present discussion is concerned.

One final bit of information comes from Mowrer. In a recent defense of two-factor theory (30), he has quoted incidental observations by Gantt to the effect that when a conditioned-response pattern embraces both motor and cardiac elements, the visceral, cardiac component often persists after the motor has disappeared. Such an observation, if well founded, might also damage an artifact theory. The original passage from Gantt goes on to say, however, that "the respiratory component also accompanies the cardiac" (9, p. 51). Evidently, then, Gantt means by "motor component" only really gross skeletal behavior. Less spectacular somatic responses not only could occur along with the autonomic responses, but admittedly do.

It begins to appear that an artifact theory is remarkably easy to defend. As a matter of fact, it is almost invulnerable. Unless an experimenter is meticulous in the elimination of skeletal responses, it will always be possible to account for visceral conditioning by postulating undetected somatic activity. And whenever visceral conditioning fails, one can always claim that no effective skeletal behavior was aroused (a stand already taken above with respect to the failure of pupillary conditioning).

This is an unfortunate situation scientifically, but it is not unique. In the current controversy over "latent learning," for instance, the reinforcementalists find themselves in exactly the same logical position. The strategy in that instance has been to attempt to eliminate all conceivable opportunities for reinforcement to operate, expecting thus to minimize "latent learning." Perhaps an analogous attack can be made upon the problem here at hand.

Related Notions

The concepts sketched above are not, of course, completely original. There has long existed a general feeling that the conditioned reflex is somehow vaguely fraudulent. Hilgard, who has displayed a lasting interest in the systematic status of conditioning (cf. 13, 14, 15, 16), has quite recently taken a position essentially antecedent to the one adopted here. Two excerpts from his 1948 book will define his stand:

> . . . experiments in which autonomic conditioning takes place (salivation, galvanic response) are full of indirect accompaniments [of a reinforcemental variety]. When the circumstances seem almost ideal for demonstrating . . . conditioning, as in attempts to condition pupillary contraction by presenting a tone along with a light, it is extremely difficult to obtain any conditioning at all. The few cases which have found conditioning are in doubt . . . (13, pp. 119–120).
>
> . . . *there is little evidence that the simultaneous or nearly simultaneous occurrence of an incidental stimulus and an unconditioned response is the sufficient condition for establishing a sensori-motor association between them* (13, p. 334).

An earlier (1935) and more general expression of somewhat the same sentiment can be found in a note by Foley:

> . . . experimental work on conditioning in complex organisms . . . is often completely vitiated by the fact that certain implicit reactions, with their attendant stimulations, are frequently occurring simultaneously or temporally adjacent to the predetermined stimulations experimentally administered by the investigator (7, p. 444).

Foley, however, does not fix his suspicions upon reinforcemental factors as plainly as does Hilgard. Neither did Freeman, who summarized his experiments of 1930, on the GSR, in much the same vein:

> Since the quantitative results were obtained in a manner similar to that of animal experimentation, I have interpreted them as "conditioned responses"; but it seems that they might just as easily be interpreted as physio-

logical accompaniments of the attitudes of expectation and surprise (8, p. 534).

Cook and Harris (6), Lazarus and McCleary (23), and Mowrer in an earlier paper (28) have expressed similar opinions on the conditioned galvanic skin response. With respect to salivary conditioning, Razran (cf. 34, 35, 36) and Zener (48) have also emphasized motivational and attitudinal factors in quite general terms.

The formulation arrived at in the present paper should not be confused with an already rather common one. A great many theorists have maintained, in one way or another, that what actually develops as a result of stimulus contiguity is an association, a cognition, an expectancy, or, perhaps, a conditioned sensation. At any rate, this general view represents essentially an extension of the contiguity principle to somatic learning (it is not always completely clear how these theorists account for the autonomic phenomena that arise during the conditioning process). Such an extension of the contiguity principle is, of course, quite opposed to the theory at hand, which sets out to obliterate the principle entirely.

The present view welcomes the suggestion that what really happens in visceral conditioning is that an "expectancy" develops. It would insist, however, that such an "expectancy" is a somatic process, inculcated by reinforcemental learning (cf. 47; 13, pp. 331–334) and occurring in the skeletal musculature, and that, as an innate byproduct of this muscular activity, the observed visceral responses arise.

REFERENCES

1. BAKER, L. E. The pupillary response conditioned to subliminal auditory stimuli. *Psychol. Monogr.*, 1938, **50**, No. 3 (Whole No. 223).
2. BEIER, D. C. Conditioned cardiovascular responses and suggestions for the treatment of cardiac neuroses. *J. exp. Psychol.*, 1940, **26**, 311–321.
3. BITTERMAN, M. E., & HOLTZMAN, W. H. Conditioning and extinction of the galvanic skin response as a function of anxiety. *J. abnorm. soc. Psychol.*, 1952, **47**, 615–623.
4. BORING, E. G. *Sensation and perception in the history of experimental psychology.* New York: D. Appleton-Century, 1942.
5. CASON, H. Conditioned pupillary reactions. *J. exp. Psychol.*, 1922, **5**, 108–146.
6. COOK, S. W., & HARRIS, R. E. The verbal conditioning of the galvanic skin reflex. *J. exp. Psychol.*, 1937, **21**, 202–210.
7. FOLEY, J. P. A critical note on certain experimental work on the conditioned response. *J. gen. Psychol.*, 1935, **12**, 443–445.
8. FREEMAN, G. L. The galvanic phenomenon and conditioned responses. *J. gen. Psychol.*, 1930, **3**, 529–539.
9. GANTT, W. H. Psychosexuality in animals. In P. H. Hoch & J. Zubin (Eds.), *Psychosexual development in health and disease.* New York: Grune & Stratton, 1949. Pp. 33–51.
10. GIRDEN, E. Generalized conditioned responses under curare and erythroidine. *J. exp. Psychol.*, 1942, **31**, 105–119.
11. GIRDEN, E., & CULLER, E. Conditioned responses in curarized striate muscle in dogs. *J. comp. Psychol.*, 1937, **23**, 261–274.
12. HARLOW, H. F., & STAGNER, R. Effect of complete striate muscle paralysis upon the learning process. *J. exp. Psychol.*, 1933, **16**, 283–294.
13. HILGARD, E. R. *Theories of learning.* New York: Appleton-Century-Crofts, 1948.
14. HILGARD, E. R., DUTTON, C. E., & HELMICK, J. S. Attempted pupillary conditioning at four stimulus intervals. *J. exp. Psychol.*, 1949, **39**, 683–689.
15. HILGARD, E. R., & MARQUIS, D. G. *Conditioning and learning.* New York: D. Appleton-Century, 1940.
16. HILGARD, E. R., MILLER, J., & OHLSON, J. A. Three attempts to secure pupillary conditioning to auditory stimuli near the absolute threshold. *J. exp. Psychol.*, 1941, **29**, 89–103.
17. HUDGINS, C. V. Conditioning and the voluntary control of the pupillary light reflex. *J. gen. Psychol.*, 1933, **8**, 3–51.

18. HUDGINS, C. V. Steckle and Renshaw on the conditioned iridic reflex: a discussion. *J. gen. Psychol.*, 1935, **12**, 208–214.

19. HULL, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.

20. KENDLER, H. H. Reflections and confessions of a reinforcement theorist. *Psychol. Rev.*, 1951, **58**, 368–374.

21. KENDLER, H. H., & UNDERWOOD, B. The role of reward in conditioning theory. *Psychol. Rev.*, 1948, **55**, 209–215.

22. KOSUPKIN, J. M., & OLMSTEAD, J. M. D. Slowing of the heart as a conditioned reflex in the rabbit. *Amer. J. Physiol.*, 1943, **139**, 550–552.

23. LAZARUS, R. S., & MCCLEARY, R. A. Autonomic discrimination without awareness: a study of subception. *Psychol. Rev.*, 1951, **58**, 113–122.

24. LINDSLEY, D. B., & SASSAMAN, W. H. Autonomic activity and brain potentials associated with "voluntary" control of the pilomotors (mm. arrectores pilorum). *J. Neurophysiol.*, 1938, **1**, 342–349.

25. MENZIES, R. Conditioned vasomotor responses in human subjects. *J. Psychol.*, 1937, **4**, 75–120.

26. MILLER, N. E. Comments on multiple-process conceptions of learning. *Psychol. Rev.*, 1951, **58**, 375–381.

27. MOORE, A. U., & MARCUSE, F. L. Salivary, cardiac and motor indices of conditioning in two sows. *J. comp. Psychol.*, 1945, **38**, 1–16.

28. MOWRER, O. H. Preparatory set (expectancy)—a determinant in motivation and learning. *Psychol. Rev.*, 1938, **45**, 62–91.

29. MOWRER, O. H. On the dual nature of learning—a re-interpretation of "conditioning" and "problem solving." *Harv. educ. Rev.*, 1947, **17**, 102–148. (Reprinted in Mowrer, O. H., *Learning theory and personality dynamics.* New York: Ronald, 1950. Pp. 222–274.)

30. MOWRER, O. H. Two-factor learning theory: summary and comment. *Psychol. Rev.*, 1951, **58**, 350–354.

31. NOTTERMAN, J. M., SCHOENFELD, W. N., & BERSH, P. J. Conditioned heart rate response in human beings during experimental anxiety. *J. comp. physiol. Psychol.*, 1952, **45**, 1–8.

32. PAVLOV, I. P. *Conditioned reflexes.* London: Oxford Univer. Press, 1927.

33. PAVLOV, I. P. The reply of a physiologist to psychologists. *Psychol. Rev.*, 1932, **39**, 91–127.

34. RAZRAN, G. H. S. Conditioned responses: an experimental study and a theoretical analysis. *Arch. Psychol.*, 1935, No. 191.

35. RAZRAN, G. H. S. Attitudinal control of human conditioning. *J. Psychol.*, 1936, **2**, 327–337.

36. RAZRAN, G. H. S. Conditioning and attitudes. *J. exp. Psychol.*, 1939, **24**, 215–226.

37. ROESSLER, R. L., & BROGDEN, W. J. Conditioned differentiation of vasoconstriction to subvocal stimuli. *Amer. J. Psychol.*, 1943, **56**, 78–86.

38. SHEFFIELD, F. D. The contiguity principle in learning theory. *Psychol. Rev.*, 1951, **58**, 362–367.

39. SKINNER, B. F. *The behavior of organisms.* New York: D. Appleton-Century, 1938.

40. SPENCE, K. W. Cognitive vs. stimulus-response theories of learning. *Psychol. Rev.*, 1950, **57**, 159–172.

41. STECKLE, L. C. Two additional attempts to condition the pupillary reflex. *J. gen. Psychol.*, 1936, **15**, 369–377.

42. STECKLE, L. C., & RENSHAW, S. An investigation of the conditioned iridic reflex. *J. gen. Psychol.*, 1934, **11**, 3–23.

43. STERN, F. An investigation of pupillary conditioning. Unpublished doctor's dissertation, Univer. of Washington, 1948.

44. THORNDIKE, E. L., *et al. The fundamentals of learning.* New York: Teachers Coll., Columbia Univer., 1932.

45. WEDELL, C. H., TAYLOR, F. V., & SKOLNICK, A. An attempt to condition the pupillary response. *J. exp. Psychol.*, 1940, **27**, 517–531.

46. WELCH, L., & KUBIS, J. The effect of anxiety on the conditioning rate and stability of the PGR. *J. Psychol.*, 1947, **23**, 83–91.

47. WOODWORTH, R. S. Reënforcement of perception. *Amer. J. Psychol.*, 1947, **60**, 119–124.

48. ZENER, K. The significance of behavior accompanying conditioned salivary secretion for theories of the conditioned response. *Amer. J. Psychol.*, 1937, **50**, 384–403.

49. ZENER, K., & MCCURDY, H. G. An analysis of motivational factors in conditioned behavior: I. The differential effect of changes in hunger upon conditioned, unconditioned, and spontaneous salivary secretion. *J. Psychol.*, 1939, **8**, 321–350.

(Received August 6, 1953)

Psychological Review
Vol. 61, No. 4, 1954

DO INTERVENING VARIABLES INTERVENE?

J. R. MAZE

University of Sydney

There is a kind of fallacy to which psychology seems especially prone (although it is not by any means peculiar to it) and which has been pointed out many times under different names —e.g., "faculty-naming," "hypostatization," "the postulation of imaginary forces," "verbal magic." But its logical structure has rarely been made explicit, and one finds that even those writers who attack it frequently commit it themselves. It seemed on first reading that MacCorquodale and Meehl, in their very provocative and valuable paper (11), had gone a good way toward clarifying this problem, but the fact that their authority has been invoked in support of such widely diverse views, including methodologies that appear particularly disposed toward hypostatization, might be taken as a sign that their paper has failed of its effect, and that a re-examination of it might be profitable.

Its main burden was to draw a distinction between two kinds of theoretical concept—"intervening variables," whose meaning and truth were completely reducible to those of the "empirical relationships" with which they dealt, or which they described, and "hypothetical constructs," which had "surplus meaning" so that the truth of statements involving them was *not* completely reducible to the truth of statements about the empirical relationships in connection with which they were hypothesized. MacCorquodale and Meehl do not contend that either of these kinds is illegitimate. What they do object to is the surreptitious use of intervening variables as if they were hypothetical constructs—as if they could sustain the functions of the latter. So far, the argument seems perfectly sound, and the fallacy to which they are pointing seems to be the "verbal magic" one, i.e., giving a name to a certain kind of event and then using that name as if it accounted for the *occurrence* of that kind of event. One example of this, I suggest, might be to use the phrase "having a valence" as meaning "being the object of our striving," and then seeming to account for our striving for something by saying that it has a valence for us.

But when one considers their criticism of "libido," one feels that they do not quite make the point that the situation one cannot use the intervening variable to account for is just the situation whose name it is. They simply say that "certain puzzling phenomena are *deduced* ("explained") by means of the various properties of libido . . ." (11, p. 105). And in their examples of "pure" intervening variables, we find some that could hardly avoid being used in the illegitimate way—e.g., "valence" itself —and some that seem quite distinct from the sort of "calculational device" (to use Spence's phrase) which, following Tolman's scheme for "breaking down into more manageable form the original complete f_1 function" between behavior and the independent experimental variables (17), they originally offer as intervening variables.

This confusion, I contend, comes about mainly because they are not clear on what the "empirical relationships" they are discussing are between, or, in general, what sort of thing can have relations. Thus, by manipulating the four variables that enter into Hull's equation for habit strength (9)—number of reinforcements, delay in reinforcement, amount of reward, and stimulus-response asynchronism (*N, t, w,* and *t'*)—they show clearly that intervening variables in the "calculational device" sense depend on quite arbitrary groupings of the empirical variables concerned, and contend that from these four they "could define 15 alternative and equivalent sets of intervening variables" (11, p. 98). But concerning one of these, a new intervening variable involving only *N,* and which they suggest might be called "cumulative reinforcement," they say: "Suppose now that a critic asks us whether our 'cumulative reinforcement' really *exists.* This amounts to asking whether we have formulated a 'correct statement' concerning the relation of this intervening variable to the anchoring (empirical) variables." Now, to ask whether *its* relation to the empirical variables is correctly stated is already to treat it as "existing," although this question of existence, which is brought in at many points throughout the essay, is not relevant since, as Bergmann (3) points out, ratios between the quantities of different variables exist even though the number by which a ratio may be expressed may not itself stand for a quantity *of* anything. More importantly, to ask about "its" relation to anything is to treat it as being the sort of thing that can be a term of a relation—that is, as being qualitative, as being some state or condition or "stuff" that there can be quantities of. Such a notion must always have "surplus meaning"; that is, a term of a relation must have some nature, some collection of properties, other than its having that relation; otherwise it would be unintelligible to say that *it* had that relation. What would "it" refer to?

Now, such a conclusion is precisely what MacCorquodale and Meehl want to avoid, but it is entailed by their speaking of "its relation to the empirical variables." What should really be considered is whether a "correct statement" of the relationship of the *number of reinforcements* to *response strength* has been formulated. Concerning $_SH_R$ in general, the point is that its "existence" is a question of whether or not the functional relationship between environmental variables and response strength has been correctly stated, which is precisely why Hull retained the *S* and the *R* in his notation: that is, in order to stress the point that it is a mathematical relationship between something done to the organism and something done by the organism, and is not in itself in any precise sense a description of the organism.

It is interesting, however, to note that Hull himself makes precisely the same error in giving his developed account (10) of what he means by the apocalyptic phrase "anchoring variables at both ends." He says: ". . . my own system . . . requires that the habit strength ($_{\dot S}H_{\dot R}$), afferent impulse ($\dot s$), and drive intensity (*D*) must each be calculable from their antecedent conditions, that the nature and magnitude of the reaction potential ($_{\dot S}E_{\dot R}$) must be calculable from the values of $_{\dot S}H_{\dot R}$, $\dot s$, and *D* taken jointly, and that the nature and magnitude of the several reaction functions . . . must each be calculable from $_{\dot S}E_{\dot R}$" (10, p. 281). But the point is that the formulae for calculating "habit strength," "drive," etc. from "their" antecedent conditions were derived in the first place from the series of experiments in which variations in each of the environmental variables listed (the others being held constant) were correlated with variations in the different measures of response strength, not in any sense with variations in "intervening variables." Thus this process of calculation and verification that Hull is describing is just a matter of verifying those empirical correlations with response strength for *different values* of the antecedent conditions—i.e., of checking the curve fitting at different points of the curve.

We have to be clear, then, that the empirically found mathematical relations are not between (for example) ${}_{S}H_{R}$ and its antecedents on the one hand, and between ${}_{S}H_{R}$ and its consequents on the other, but just between the antecedent and consequent events—in fact, that that relationship is just what ${}_{S}H_{R}$ *is*. The other notion of what it is—i.e., some qualitative condition of the organism, a given amount of it being produced by specified environmental events, and in its turn producing a given strength or frequency of response—is what MacCorquodale and Meehl refer to as a hypothetical construct; but, as I suggested, their description of the meaning of intervening variables seems in some places to imply that sort of notion.

Now that might seem hardly more than a slip of the pen on their part, especially since they go on to say (11, p. 99) that "when habit strength *means* the product of the four functions of w, t, t', and N, then if the response strength is related to these empirical variables in the way described, habit strength 'exists' in the trivial sense that the law holds"—although one cannot see why this should be thought trivial if nothing else was ever expected of ${}_{S}H_{R}$. But if that really is what MacCorquodale and Meehl mean by "intervening variable" then why do they not press home their sketched-in criticism of any rigid hierarchy of intervening variables? They suggest that Tolman's argument (17) that it is easier to determine the complicated f_1 function by parts than "as a whole" is "not very cogent." They do not elaborate this point, but I take it that what they are suggesting is that one can only do it step by step, so that setting up a doctrine which advocates the use of such intermediate steps cannot be any contribution (if that is really all it offers). One might say that if we want to discover the product of 3, 4, and 5, then it is not really possible to multiply them all together at once; we have to find the product of, say, 3 and 4 and then multiply that product by 5. But that procedure has no advantage whatever over first taking 3 and 5 together, or 4 and 5, so long as we really are concerned only with convenience of computation. In their manipulation of w, t, t', and N, MacCorquodale and Meehl provide all the materials necessary for showing that the same applies to the grouping of *all* the environmental variables studied by Hull, including as well as those four the maintenance and stimulus variables effective at the time of any given trial. If convenience of manipulation is the only concern, it must be very difficult to show why that particular four out of all these factors should be taken together, and why their product should be given a special name.

Hull, too, feels sensitive on this point, since he puts a similar objection into the mouth of "his friend Woodrow" (10, p. 284). His replies to it are quite unconvincing: he says in the first place that even if we did put all the subordinate equations together to form one, it would be found still to contain "in some form or other the mathematical equivalents of the various equations linking the observable and the hypothetical unobservable elements of the situation." But to say that the groupings are merely for convenience is to deny that one is thinking of any "hypothetical unobservable elements," and as for the rest, to say that the mathematical representatives of the *observable* elements (and the relations between them) would still be found in the master equation is to admit the point that "Woodrow" is making.

Hull's second point in reply to the criticism gives up the attempt to justify intervening variables on a formal basis and introduces the material consideration that the action of the four training variables might be temporally remote from that of the critical stimulus and the state of need. And "while it is perfectly possible to put into a single equation the values of events which occur at very different times," still those past events cannot be causally active now, and "$_{\dot{S}}H_{\dot{R}}$ is merely a quantitative representation of the perseverative after-effects" of those four training variables (10, p. 285). Hull, then, groups the variables in this regular way not merely as a matter of convenience in calculation (if indeed one way could be more convenient than another), but also because he regards the variables in any group as acting together to build up some specific condition in the animal, and the intervening variable based on that group would then be thought of, if not actually as a "measure" of that condition, at least as varying quantitatively in direct relation with it.

One might contend, then, that some such speculation about the accompanying qualitative states could be the only reason for clinging to a set order of groups and giving names to their products when these products have still to be combined to discover the probability of a given response. But as "calculational devices," the only solid content for the notion of "intervening variables" is just the collection of correlation coefficients *between* response strength and each of the empirical variables found relevant to it. Without some such mathematical form we could not give even this slight meaning to the term "intervening," since we would be left only with the assertion that a given factor has some causal connection with a given response, and causality does not in any sense *intervene between* antecedents and consequents—it is just the antecedent *event* that produces the consequent, not "causality."

The finding of such relationships, by the way, indicates the only sound meaning for the phrases "empirical variables" and "empirical relationships"—that is, what is really meant is more like "empirically-*found* relationships," and even here, for the empiricist, the phrase is redundant, since there is no other way of arriving at knowledge than by finding it empirically. By using the word in this way, MacCorquodale and Meehl convey a vague suggestion that empirical relationships are to be compared, unfavorably, with more certain, more fundamental, more intelligible "laws," and it is strange that this hint of the rational is to be found amongst the positivists at large (e.g., 7).

One point that helps to preserve this distinction in thought is that it is very difficult to discover perfect regularity in the sense of coextension when we are looking for causal connections. Frequently one can find only conditions that are sufficient but not necessary, or necessary but not sufficient. This has also led to the current suspicion that there must be something wrong with causality, has led to the not-talking about it, and to the being content with (and compensatory exaltation of) correlation coefficients and statements of probability in general. Without denying the usefulness of such mathematical procedures as a first approach to a confused field, one can still say that the rejection of indeterminism will involve denying the adequacy of probabilities in science, and will involve affirming the presence of a criterion for every case in which a necessary condition is not also sufficient, and vice versa. That is to say, it is always theoretically possible to extend our knowledge of the effective conditions until we arrive at a set that is necessary *and* sufficient for any given event. In seeking causes, such criteria can only be found by considering *both* the nature of the thing in which the change is produced and the nature of the thing acting on it (1). A recognition of this principle seems characteristic of the work of those scientists who increase the number of *general* propositions known in their subjects. To take only two examples in psychology, there is the recent work of Tinbergen (16) with his emphasis on the need for both a specific type of stimulus situation and a specific bodily condition for the production of an instinctive response; and there are the explicit formulations of Freud (8), which might have prevented a good deal of the dreary heredity-environment controversy if they had been better known. Watsonian behaviorism, then, was based on a false premise—". . . given the response the stimuli can be predicted; given the stimuli the response can be predicted" (18, p. 167)—since the same stimulus situation will produce different responses in the same animal according to changing conditions in the animal, a point recognized in part by Hull in his emphasis on *D*.

If psychology is ever to make predictions, then, rather than mere statements of probability, it is, in MacCorquodale and Meehl's terms, committed to "hypothetical constructs" or, more precisely, to making hypotheses (and trying to verify those hypotheses) about what processes in the animal mediate a given change in its behavior. As Bergmann points out (4), these hypotheses will be assertions that processes of a kind which we have known in other places are going on (so far unobserved) in this place—the point being that we can arrive at the notion of any term or kind only by confronting such a kind, only from experience. Where else could we get the material for our fantasies? Wherever we seem to "construct" the notion of a kind, it is always by conjoining properties that we have encountered (separately) in actual things—and if that is so, then it is sufficient to dispose of the doctrine that there *can be* "convenient fictions" in science, when these are said to be neither true nor false. MacCorquodale and Meehl themselves reject this latter doctrine, but their retention of the phrase "hypothetical constructs" renders them liable to be misunderstood as supporting it (cf. Bergmann, 3), as does their question whether constructs are "existential" —as though it were a real possibility that we could talk about some *thing* which was not existential.

Now, "intervening variables" in the sense described above—as correlations between events impinging on the organism and responses produced by the organism—cannot be states or properties or qualities of the organism itself, even though they presuppose the existence of such qualities. The environmental variables may have been grouped in a specific way (as Hull's are) because of some tentative speculation about a specific change produced in the animal by each group, but even so, and even if these speculations are put forward, the mathematical relations remain distinct from the hypothesized processes which are held to account for them. This is a point overlooked by MacCorquodale and Meehl when they say (11, p. 101) that "there are various places in Hull's *Principles* where the verbal accompaniment of a concept, which in its mathematical form is an intervening variable in the strict (Tolman) sense, makes it a hypothetical construct." That makes it seem that they always have half thought of intervening variables as being in some very vague sense "in" or "of" the organism (and have accepted them thus as legitimate), the only point being that they are not to be ascribed properties, not to be characterized in any way (cf. O'Neil, 12), else they become hypothetical constructs. This suggestion is borne out in their discussion of Skinner's treatment of emotion as "a 'state of the organism' which alters the proportionality between reserve and strength." They say (p. 102): "The 'state' of emotion is not to be described in any way except by specifying (a) The class of stimuli which are able to produce it and (b) The effects upon response strength. Hence emotion for Skinner is a true intervening variable, in Tolman's original sense." Now, previously they had described "state" as a "wholly noncommittal word" which specifies "nothing except that the conditions are internal" (p. 97). But this point about internality is precisely the one at issue in discussing the fallacy of hypostatization. We may agree that there always will be some condition of the organism in virtue of which, in specified circumstances, a given response will be produced, but, in direct contradiction to MacCorquodale and Meehl, it *must* be described in ways other than its relations to its antecedents and its consequences (or in general its relations to anything); otherwise (*vide supra*) statements about what produced *it* and what *it* produced (or about any of its relationships) will not be intelligible. In saying that it is not to be described except by specifying those relations, and in saying (by calling it a "true intervening variable") that it is identical with the "empirical relations" and yet is a state of the organism, MacCorquodale and Meehl are setting up the notion of something whose whole nature it is to stand in a given relationship. Even if it is objected that it is the organism that has the relationships, and that they are not its whole nature (since it has many other properties) but only a part of it, still this modified doctrine faces the same difficulty—namely, that it is strictly "unspeakable" (2) since we can only grasp a relationship if we can distinguish the terms that have it to each other; that is, seeing them as distinct, as having distinct natures, is a part of seeing them as related. If we say that its relationship to a certain stimulus situation is part of the organism's nature, then the whole relation (including its other term, the stimulus) seems to be brought *within* the organism, so that we cannot really understand the assertion that there *is* this relationship between distinct terms (this being the insuperable problem for Lewin's life-space).

This unworkable view that a thing can be in whole or in part made up of its relationships is the crux of hypostatization—of all doctrines of unseen forces or magical entities. This may be seen more clearly if we consider the organism's relations to its own responses. For the most part, a response really stands logically as a new property of the organism, i.e., as a change in its nature, and so of course it *is* a part of the organism. But we are concerned with the relationship of this new property to the state of the organism immediately *prior to* its appearance, and if we make that relationship a part of the preceding state of the organism, then we have the characteristic form of the fallacious doctrines we are discussing; that is, the organism produces that response simply because it is *in its nature* to act in that way—it has a propensity to do so. At one stroke we are absolved from seeking for those actual states of the animal which determined that, under some specific stimulation, it would produce that response, and from discovering precisely what the stimulus is, since it would seem merely an "inclining" cause at most, being in fact a necessary but not sufficient condition.

This is the very fallacy MacCorquodale and Meehl are criticizing in their attack on the use of libido to explain features of behavior, but they mistakenly think it appears when libido is ascribed properties of its own, and that the way to avoid the fallacy is carefully to keep its nature devoid of any surplus meaning, i.e., surplus to "its" functions. The same prescription seems to be the one central to operationism, and though in the long run it goes astray, it is possible to see some force in it in one specific connection—namely, in the definition of dispositional concepts, with which in one place MacCorquodale and Meehl identify intervening variables as they conceive them. We must sympathize with at least the policy implied in Stevens' cry: "Only then shall we never think of energy or consciousness as a substance . . ." (15, p. 330).

Taking "solubility in water" as a dispositional concept, then following Bergmann's argument (3, p. 98), a proper (positivistic) definition of it would be to say, e.g., that "x is soluble in water" means "if x is put in water, it dissolves." The verification of the first sentence is held to be completely reducible to that of the second, which implies (and Bergmann makes this quite explicit) that the words in the first sentence mean precisely the same as the words in the second sentence. (Some positivists hold that this complete reducibility applies to *all* defined concepts, though that extreme view seems to have been modified by Carnap [5, p. 464, ff.]. The general proposition that all definitions are nominal comes down to saying that there is no such thing as coextension, which, as I suggested above, is tantamount to a rejection of determinism.) Now, it seems to me that the suggested definition is deficient in that it does not take into account objects which in fact would dissolve in water but which never are placed in water. (Modern symbolic logic would hold that the defining sentence could be converted to "either x is not put in water *or* x dissolves," and that if the first of these disjuncts is satisfied by x never being put in water, then the *whole sentence* is true of x—i.e., the conditions for x being soluble are satisfied, and the definition is held in this way to be adequate even for the negative instances. But if we admit that meaning of "either . . . or . . . ," then, as Carnap points out [5, p. 440], the definition would include not only lumps of sugar that are not put in water, but also such things as a wooden match that is not put in water—anything, in fact, that does not meet that fate would "satisfy the conditions" for being soluble in water.)

The difficulty can be met, however, by amending the definition to read "x is *of such a nature* that if it is put in water, it dissolves," even though we do not know what that "nature," that common quality or character of soluble things, may be. Now, in principle, the verification of the presence of this character is not completely reducible to the observation that when x is put in water it dissolves, since any quality is theoretically observable, and its presence then could be established (if we knew what it was) by direct observation, without the necessity for observing its effects. But one might say roughly that the verification of solubility is thus reducible because (and this in my view is the point the positivists are really getting at here) it is not the character in question. To call a thing "soluble" is just to say that it has some unspecified character in virtue of which specified events produce a specified change in it, and that it retains that character even when those events do not materialize. In itself it merely poses the question, what this character is, but it is very frequently used as if it were the answer to that question—as if it were that character itself—and that, in fact, is the typical way in which hypostatization occurs.

Ironically, however, by its very insistence that dispositional concepts are not "substances" and do not have "surplus meaning," positivism sometimes leads to precisely the sort of mysticism it is trying to make impossible. That is, it is taken up wrongly (even by many positivists) as suggesting that there is *no* characteristic there in virtue of which the events in question take place, that the mysticism lies in going on to look for it, that the scientific procedure is to be content with "solubility," and that we cannot fall into confusion as long as we rigidly exclude from our thinking any suggestion of quality or "substance" in the matter at all (failing to see that it is only from the notion of "solubility" itself that it must be excluded). But such a course of thought (which seems to be MacCorquodale and Meehl's) makes the fallacy inevitable; if there is no relevant quality there, nothing which produces the dissolving but is describable in terms which *make no reference to* producing that effect, then that relationship (to dissolving) must be thought of as just "being in the nature of the thing." Not only is such a notion not itself a solution, but while it is retained it specifically makes a solution impossible.

The reply might be made in defense of MacCorquodale and Meehl that they plainly do recognize the possibility of finding the qualitative processes mediating any response since they recognize "hypothetical constructs" as scientifically legitimate. But in my opinion the error remains; for them it is *not* that the hypothetical construct is found alongside of, and mediating, the intervening variable—not that the qualities which determine that given events produce given consequences still remain distinct from that relation between antecedents and consequences. Rather it is that "the existence propositions . . . automatically make the construct 'hypothetical' rather than 'abstractive' " (11, p. 99)—i.e., the intervening variable *becomes* a hypothetical construct in the ascription *to it* of qualitative content. This is a further indication of what is made plain in their discussion of Skinner's "emotion": that the true intervening variable is thought of as a "state" *internal to* the organism, in some sense part of its constitution yet stripped of all qualitative content—and in that case it can only be the relativistic, mystical sort of notion that they are confusedly setting out to attack.

Although I said that there is one real set of facts indicated by some of MacCorquodale and Meehl's uses of the term "intervening variable"—namely, the mathematical relations between types of event impinging on the organism and types of response produced by it—it does not seem to me that the term "intervening variable" is necessary or suitable for referring to it, partly because of other explicit meanings that have accrued to it, and partly because the words themselves inevitably suggest some state-like thing that *intervenes between* stimulus and response. A relation is not "between" its terms even in the most neutral way; one should say, rather, that they have, or stand in, that relation.

The mere giving of a name or symbol to those relations strengthens the ever-present temptation to slip into the "imaginary force" way of thinking, especially when the correlations are less than unity. In this case we cannot say that the stimulus in question produces the specific response (because there are some cases in which it does not), but if we remain convinced that there is some connection, then we are likely to say that the stimulus results in a tendency to produce that response. But this tendency regularly finds its way inside the organism (because of a confused recognition that the state of the organism has something to do with it), and appears as a "demand" or "propensity" (or any of the multifarious species of "tendency") to make that response. (As "valence" it has found its way into the stimulus-object—i.e., the stimulus has "a tendency" to produce the response.)

Now, the use of "intervening variables" is sometimes defended by insisting that they *are* "imaginary" forces and are never intended as anything else. Their sole function is to help us grasp the observed facts and to organize our thoughts about them. But it is possible for our thoughts to be "organized" in a way that is mistaken or even meaningless. It may for the most part be true that the actual verbal forms are nothing but ways of expressing the observed connections, but the mere fact of offering "intervening variable" statements makes it vaguely seem that our knowledge is being extended, and that we are "getting to understand the facts better" in the sense of seeing how they are produced. The attribution of events to occult forces is rarely explicit because then it is so blatantly unscientific; it creeps in unacknowledged and gains its influence by default, as it were—by our failing to look for the *actual* causes.

To discover causal relations (which are always coextensive relations) we must take into account not only the nature of the forces acting on the thing in which the changes are produced, but also the properties, especially the fluctuating ones, of that thing itself (which in psychology will, for the most part, be, of course, the organism). When we discover which of its *actual* properties are involved in any given reaction, then the need to fall back on imaginary forces whose sole function is to produce those effects (i.e., on "intervening variables") will have disappeared.

REFERENCES

1. Anderson, J. Realism and some of its critics. *Aust. J. Psychol. Phil.*, 1930, 8, 113–134.
2. Anderson, J. The problem of causality. *Aust. J. Psychol. Phil.*, 1938, **16**, 127–142.
3. Bergmann, G. The logic of psychological concepts. *Phil. Sci.*, 1950, 18, 93–110.
4. Bergmann, G. Theoretical psychology. *Annu. Rev. Psychol.*, 1953, 4, 435–458.
5. Carnap, R. Testability and meaning: I–III. *Phil. Sci.*, 1936, 3, 419–471.
6. Carnap, R. Testability and meaning: IV. *Phil. Sci.*, 1937, 4, 1–40.
7. Feigl, H. Operationism and scientific method. *Psychol. Rev.*, 1945, **52**, 250–259.
8. Freud, S. Heredity and the etiology of the neuroses. In E. Jones (Ed.), *Collected Papers*. Vol. I. London: Hogarth, 1924. Pp. 138–154.
9. Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
10. Hull, C. L. The problem of intervening variables in molar behavior theory. *Psychol. Rev.*, 1943, **50**, 273–291.
11. MacCorquodale, K., & Meehl, P. E. On a distinction between hypothetical constructs and intervening variables. *Psychol. Rev.*, 1948, **55**, 95–107.
12. O'Neil, W. M. Hypothetical terms and relations in psychological theorising. *Brit. J. Psychol.*, 1953, **44**, 211–220.
13. Skinner, B. F. *Behavior of organisms.* New York: Appleton-Century, 1938.
14. Spence, K. W. The postulates and methods of 'behaviorism.' *Psychol. Rev.*, 1948, **55**, 67–78.
15. Stevens, S. S. The operational basis of psychology. *Amer. J. Psychol.*, 1935, **47**, 323–330.
16. Tinbergen, N. *The study of instinct.* London: Oxford Univer. Press, 1951.
17. Tolman, E. C. The determiners of behavior at a choice point. *Psychol. Rev.*, 1938, **45**, 1–41.
18. Watson, J. B. Psychology as the behaviorist views it. *Psychol. Rev.*, 1913, **20**, 158–177.

(Received August 31, 1953)

Psychological Review
Vol. 61, No. 4, 1954

RESPONSE FACTORS IN HUMAN LEARNING[1]

GEORGE MANDLER

Yale University[2]

Theoretical treatments of human learning phenomena have been largely confined to an analysis of stimulus variables. In recent years, however, more attention is being paid to response factors, especially in the investigations of Underwood (28), Morgan and Underwood (20), and Osgood (22). This paper will present a theoretical framework, applicable to human learning and thinking problems, which will stress response factors. Attention will also be paid to the point, made by McGeoch (18) and others, that most "new" learning in human adults is at least partly a transfer phenomenon. The paper will be particularly concerned with the differentiation of stimulus conditions depending on the evocation of responses made by the organism, the relationship between overt and symbolic responses, and the transfer and overlearning of these responses. Thus, stimulus factors will be viewed as dependent upon the particular response repertory and previous experiences of the individual. This is in line with the position of Sperry (27) who has recently argued, from a neurological point of view, for a response approach to problems of perception and thinking.

The definitions and assumptions which represent the theoretical structure will be specified and some applications will be given, with particular reference to the effect of overlearning on transfer. The reader will note that some of the assumptions are deductive corollaries of others, or are related to other theoretical systems. For the present purposes they will be stated as assumptions.

DEFINITIONS

Stimulus. The term stimulus will be used as defined by Hull (15), essentially in terms of receptor input. Hull defined as "actual stimuli" those events which activate a receptor.

Overt response. This term will refer to any observable activity of the organism.

Symbolic response. A human organism will be considered to have made a symbolic response analogous to the overt response if he reports the perception of the overt response without performing the overt response.

Reinforcement. The term reinforcement will refer to an event in the environment which indicates to an individual the correct performance of a response. (No position is taken in this paper as to the specific mechanism of reinforcement. Presumably the above formulation is consistent with all current theories of reinforcement.)

ASSUMPTIONS

The Differentiating Response

1. A stimulus is differentiated from other stimuli when it evokes a response different from the response evoked by other stimuli. This differentiating response will be designated as R_S. The

[1] This paper is part of a dissertation submitted to the Graduate School, Yale University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The writer wishes to thank Drs. F. D. Sheffield, C. I. Hovland, and N. E. Miller for their suggestions and help in the preparation of this paper.

[2] Now at Department of Social Relations, Harvard University.

R_S can belong to any class of responses, i.e., it can be verbal, motor, or symbolic, depending on the original learning experiences of the individual.

2. When identical R_S responses have been frequently reinforced for two or more different stimuli, these stimuli, other things being equal, will be perceived as identical. Conversely, when different R_S responses have been reinforced to two or more different stimuli, these stimuli will be perceived as different. Identity or difference of stimuli, qua stimuli, depends solely on receptor stimulation, but identity or difference in the perception of these stimuli depends upon the differentiating responses. Thus identical stimuli cannot be perceived as different, but different stimuli can be perceived as identical. Another important implication is that two stimuli which differ in receptor stimulation, but do not evoke any differentiating responses, cannot be perceived as different.

3. Several different differential responses can be associated with any one stimulus and, other things being equal, they will differ only in terms of the probability of their evocation, which is a function of their reinforcement history, as in Hull's "habit family hierarchy" (14).

It will be noted that the present concept of differentiating responses is closely related to Dollard and Miller's (7) concept of cue-producing responses. Their description of attaching the same (or distinctive) cue-producing responses to distinctive (or similar) stimulus objects does not differ from the present treatment. The present statement does imply, however, that stimuli cannot be differentiated unless different responses are evoked.

The adult human organism responds to complex stimulus situations with a variety of previously learned responses which in turn become the stimuli in the learning of new responses, i.e., they "mediate" the new associations. Thus, in a concept-formation experiment, the learning of a nonsense syllable response to all "green" stimuli is an example of mediated learning. On the other hand, a child's learning to differentiate the colors "red" and "green" (different stimuli) by differentially attaching the two verbal responses would be considered original learning.

Response Integration

1. Many responses performed by human organisms consist of aggregates of several subresponses which may be innate or acquired.

2. With successive repetitions of a response aggregate, the separate responses eventually become stimuli for each other such that any part of the response aggregate will tend to evoke the whole response aggregate. This process will be referred to as integration of the response.[3]

3. Integration is an increasing function of reinforced repetitions of the response aggregate.

4. The growth of integration is dependent upon the elimination of responses which prevent or delay reinforcement (as in "anticipatory errors").

5. The integration or association of two responses proceeds more rapidly than the association of a response with a stimulus. Thus, it is easier to learn a new response to a stimulus which already evokes a differentiating response than to a new unfamiliar stimulus.

6. The integration of a response aggregate proceeds more rapidly when the response units belong to the same effector modality. Differences in effector

[3] Integration as used here does not refer to the simple chaining of responses, but rather to the simultaneous elicitation of an aggregate of responses. Responses that occur as overt chains very often are integrated, in the present sense, at the symbolic, perceptual level.

modality refer to differences in effector organs utilized in making the response. Thus it would be easier to integrate two verbal responses than a verbal and a motor response.

This conception of integration is closely related to Hollingworth's (11) concept of redintegration. Guthrie (9) has modified Hollingworth's concept in terms of parts of a response tending to condition each other.

Symbolic Responses

1. Any overt response which is perceived by a human organism evokes a symbolic response analogous to the overt response.

2. The symbolic response tends to be activated whenever the overt response is performed. Evocation of the symbolic response, however, tends to elicit the overt R_S only if motivation to perform the response is present.

3. Whenever a stimulus evokes two separate integrated responses, the two symbolic analogues may also be activated and associated so that, on future presentations of a stimulus which evokes only one of the responses, both symbolic responses will be activated.

4. Symbolic responses can be associated with other symbolic or nonsymbolic responses. In particular, they can be associated with verbal responses.

5. Previously learned verbal responses can have inhibiting effects which prevent occurrence of an overt response. When a particular overt response no longer leads to reinforcement, verbal statements as to its incorrectness can be attached by this experience to the symbolic analogue. On future occasions when the symbolic response occurs anticipatory to performing the overt response, the inhibiting verbal response can effectively forestall the error.

6. In a learning task, the symbolic analogue of a response aggregate will differ from one trial to the next as long as irrelevant overt responses are still present and errors are still made. Thus, irrelevant responses which do not prevent, but are also not correlated with, reinforcement will from time to time be represented in the symbolic analogue. Necessary overt responses will continue to be represented and reinforced, while the irrelevant responses will drop out. In particular, the symbolic analogue is expected to be most distinctive and constant after many contiguous repetitions of the same response aggregate, i.e., after errors have been eliminated and overt performance has reached an asymptote.

The actual modality of the symbolic response is not relevant to the applicability of this concept. It appears advisable to leave stipulation of modality unspecified, which makes it possible to extend the concept of symbolic responses to both verbal and nonverbal behavior.

Applications

In this section an attempt will be made to integrate, in terms of the present theoretical framework, some of the empirical results that have been obtained in studies of human learning. The studies quoted are intended to be representative of the empirical data in a particular area rather than exhaustive listings of these data.[4]

Learning of Differentiating Responses

It is to be expected that complex stimuli will evoke several differentiating responses previously learned to different parts of the stimulus pattern. At the same time, however, differentiating responses can be learned to the total stim-

[4] In the present discussion, no statements have been made about the effects of differential motivation and related concepts. Such phenomena are presumably operative in addition to the effects discussed here.

ulus. Rossman and Goss (24) found that, while subjects used recently acquired verbal differentiating responses to distinguish stimuli in new paired-associates tasks, they looked for "identifying parts" of stimulus terms more frequently and found these more helpful. We would expect such a preference since highly overlearned responses to the identifying parts presumably have a higher degree of probability of evocation than recently learned nonsense syllables.

On the other hand, adult human behavior also provides many examples of the association of identical differentiating responses to different stimuli. Thus, the same word written in several different handwritings or in print will evoke the identical response and will be perceived as identical unless differentiating responses referring to the stimulus differences are actually evoked.

Differentiating responses will be associated not only with the experimentally controlled stimulus but also with other aspects of the total situation. The integrated response is reinforced in the context of the experimental situation, particularly in regard to the experimenter, the apparatus, and so forth. A recent study by Bilodeau and Schlosberg (3) showed more retroactive inhibition when the experimental room was the same for both tasks than when the interfering task was learned in a different room.

Prior training with attendant integration of a differentiating response is expected to shorten the learning process on future occasions when that response is used. Hovland and Kurtz (12) found that prior familiarization with nonsense syllables facilitated learning of lists using those nonsense syllables.

Concept Formation

Under this heading will be discussed only those aspects of so-called concept formation in which a subject is required to learn a common response to a class of stimuli. The experimental situation has most frequently been exemplified by Heidbreder's experiments (10).

It is assumed that in the presentation of a number of stimuli, many differentiating responses will be evoked by each stimulus. The stimuli will have been predifferentiated to varying degrees, and different aspects of each stimulus will evoke different responses. In most cases these responses are verbal responses which also elicit their symbolic analogues. At the same time, the subject learns to make the new paired-associates response (nonsense syllables in Heidbreder's experiments). These new responses, which have never before been evoked in the presence of these particular stimuli, are now associated with the already learned symbolic responses. In the process of successive presentations, the symbolic response which corresponds to the "concept" will be associated more frequently than any other with the new response. Thus, if one of the prior differentiating responses is "face," then the new response (the nonsense word) will be associated with the symbolic analogue "face." In successive presentations of instances of this concept, this association will be reinforced so that, eventually, the evocation of the symbolic response "face" will also evoke the new *name* of the concept—the correct nonsense response. Figure 1 shows a schematic representation of this process; R_{S_x} is the symbolic response common to the stimuli, and will be more frequently associated with the concept response than other differentiating responses. This is similar to Hull's (13) theoretical description of concept formation. The difference is that, in the present formulation, the concept response is (at least at first) associated with the respective differentiating responses rather than directly with the stimulus components.

A recent study by Baum (2) indicates

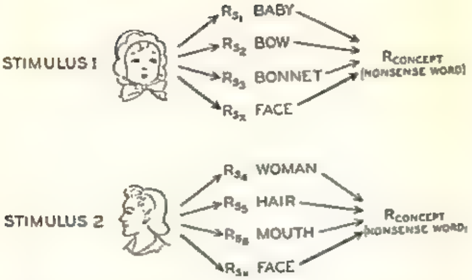

FIG. 1. Two instances of the same concept (Stimuli 1 and 2) eliciting one common ("face"), and several different, differentiating responses

that ease of concept attainment is a function of the discriminability of the stimuli. Her findings imply that degree of previous learning of the differentiating response determines ease of concept learning. In the extreme case, it would be expected that a stimulus which evokes no previously learned differentiating responses—the completely "unfamiliar" stimulus—could not be one of a class of "similar" stimuli. On the other hand, stimuli which have been maximally differentiated, i.e., with a high probability of evocation of a differentiating response, provide highly integrated responses, potent symbolic responses, and easy association with the new response.

Response Generalization

The general statement of this phenomenon usually implies that the learning of a response to a particular stimulus will facilitate the learning of similar responses to the same stimulus. This similarity can be specified in two dimensions. It can be either a similarity of overt parts of the two responses or a similarity of symbolic responses, as in the use of synonyms in Morgan and Underwood's experiment (20). An attempt will be made to show that response integration and symbolic responses are sufficient to explain the phenomena usually described as response generalization.

Similarity in terms of elements involves a communality of some of the parts of an integrated response. Substituting a new unit for one of the original units of the integrated response does not affect the integration of the units which are common to both aggregates. If it is assumed that the replaced unit can be dropped out fairly efficiently, this situation should produce faster learning than one in which all the units have to be integrated.

Similarity of symbolic responses, i.e., in the meaning realm, is comparable to the concept formation situation. The two synonymous responses both evoke a common symbolic response. As a concrete example, two of the synonyms used by Morgan and Underwood (20) were "dirty" and "unclean." To the extent that these two responses are associated with a common symbolic concept such as "filth," paired associate learning of the second response will be mediated by the common concept. Figure 2 diagrams this process. The common symbolic response, however, need not be as specific as the one used in the example, and it may be nonverbal. Morgan and Underwood also found that different degrees of synonymity are reflected in the degree of response generalization. The position taken here is that synonymity is a function of common symbolic representation of the two differentiating responses, which in turn would affect response generalization as found. A similar point of view, describing meaning as a function of commonly associated re-

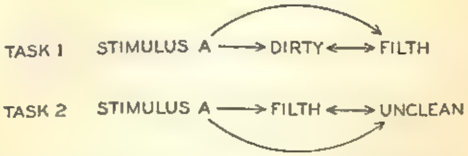

FIG. 2. The mediating and facilitating effect of a common symbolic concept in the response generalization of two synonymous responses

suggest that, with increasing training on the original task, the number of correct responses on the reversal transfer task increases (response facilitation through transfer of a previously integrated response), but that specific errors also increase (interference due to previous reinforcement of the now incorrect response).

2. *Learning to make an old response to a new stimulus.* In this condition, the primary factor is assumed to be the integration of the response, i.e., the subject learns how to perform the response. Since the response has been learned not only in the presence of its original stimulus, but also is associated with all other constant aspects of the experimental situation, the probability of its evocation is high, and once evoked and reinforced in the transfer situation, the rate of learning of the new association will be partly a function of the integration of the response and its association with the context cues. It should be pointed out, however, that measurable transfer will reach a maximum before response integration does. For example, if the new association is learned on the first or second trial of the transfer task, increased integration of the response on the original task will show little measurable transfer effects. The prediction in this situation would be increasing positive transfer as a function of correct repetitions of the response on the original task.

The earliest relevant study is that of Bair (1), whose subjects learned to press specific colored typewriter keys to color stimuli. When a new list of stimuli was presented, with the responses remaining identical to the ones previously used, he obtained positive transfer. Bruce's data (4), again corrected for warm-up effects, show increasing positive transfer as degree of original learning is increased. Bunch and Winston (5) obtained clear positive transfer effects when a previously learned nonsense syllable response had to be learned to a new stimulus.

The paucity of data in the literature relevant to this problem is compensated for by their unequivocal direction. However, one important implication from the present position has been given little experimental verification: if the response on the original task is already highly integrated (e.g., in the use of adjectives), then transfer effects should be minimal. In other words, amount of positive transfer in this situation is partly a function of degree of integration of the response at the beginning of the original task. The evidence that is available, however, is confirmatory (25).

3. *Learning to make an old response to an old stimulus when these have not been previously paired.* Three factors influence learning in the transfer task:

a. The integration of the response in the original task: facilitating effect.

b. The probability of evocation of the response, now incorrect, learned in the original task: interfering effect.

c. The symbolic "inhibition" of that incorrect response: counteracts the interfering effect of *b*.

At low degrees of learning, up to maximal strength of the original association, we would predict variable positive or negative transfer effects, depending on the relative strength of the facilitating and interfering effects *a* and *b*. When the symbolic analogue has become stable (at high degrees of overlearning), there would be an increasing tendency toward positive transfer as the interfering effect of *b* is counteracted by *c*.

One additional factor, however, is important in this particular condition. The differentiating responses evoked by the stimulus may be integrated to a greater or lesser degree with the response learned in the original task. In Conditions 1 and 2, this factor is presumably of minor importance. In Condition 1, the old response is never reinforced in the transfer situation, and the integration would be additive to the general interference effect so that like-modality responses would lead to greater initial interference. In Condition 2, the effect might delay the increasing positive facilitation since part of the integrated response (the previously correct differentiating response) has to be eliminated. In Condition 3, however, these two effects would not only be additive, but the constant re-evocation of parts of this integration would interfere with the elicitation of either part alone. A recent study by Porter and Duncan (23) has shown greater negative transfer in Condition 3 than in Condition 1, when the stimulus and response elements were both verbal, which would favor integration of stimulus and response elements. In their discussion, the authors point to the possibility of the response re-evoking the stimulus and thus leading to greater interference.

If the differentiating response is verbal and the newly learned response (in the original task) motor (i.e., if the two responses belong to different effector modalities), we would predict that the integration of these two components would be minimal and would show less negative transfer than Condition 1. Siipola and Israel (26) have presented data bearing on this latter expectation. Subjects were pretrained to learn a series of responses on telegraphic keys. They were then presented with the original task in which these codes had to be associated with letters of the alphabet. In the transfer task, the same stimuli and responses were used, but their combinations were changed. The data, measuring transfer as a function of training on the original task, show slight initial negative transfer followed by a large positive transfer effect.

Kline's (16) subjects paired authors' names with book titles. He found that, with greater degrees of prior knowledge of the *correct* authors' names, paired-associates learning was easier even when the correct response was to give wrong authors' names to the book titles. His evidence for decreased interference as a function of increasing familiarity is consistent with our prediction.

CONCLUSION

Further empirical verification of the above predictions should precede the application of this theoretical framework to more complex problems. The general emphasis on the model presented here has been on the importance of response factors in activities of the human organism prior to its introduction to an experimental situation. Phenomena such as stimulus discriminability are assumed to be a function of such previous experiences. In these terms, the differentiation of stimuli varies from individual to individual, and any general description of the discriminability of a stimulus only refers to the communality of experiences a group of individuals has had in learning differentiating responses to that stimulus. Thus, statements about stimuli, other than those referring to receptor stimulation, are limited to common social and learning experiences of subjects. In the past, studies such as Gibson's (8) have used stimuli and responses (with relatively homogeneous groups of subjects) which were most likely to result in similar differentiating learning experiences.

In reference to transfer effects, Guthrie's warning (9) that transfer is specific and not general has been taken account of. Particular attention has been paid to the fact that most human activities involve highly overlearned responses and response aggregates, and to the relevance of this phenomenon to transfer effects. If the predictions made concerning the differential effects of a subject's experiences with the stimuli and responses are borne out, then such long-accepted generalizations as Wylie's

(30), that "the transfer effect is positive when an old response can be transferred to a new stimulus, but negative when a new response is required to an old stimulus," need re-examination.

REFERENCES

1. Bair, J. H. The practice curve: a study in the formation of habits. *Psychol. Monogr.*, 1902, **5**, No. 2 (Whole No. 19).
2. Baum, Marian H. A study in concept attainment and verbal learning. Unpublished doctor's dissertation, Yale Univer., 1951.
3. Bilodeau, Ina M., & Schlosberg, H. Similarity in stimulating conditions as a variable in retroactive inhibition. *J. exp. Psychol.*, 1951, **41**, 199–204.
4. Bruce, R. W. Conditions of transfer of training. *J. exp. Psychol.*, 1933, **16**, 343–361.
5. Bunch, M. E., & Winston, M. M. The relationship between the character of the transfer and retroactive inhibition. *Amer. J. Psychol.*, 1936, **48**, 598–608.
6. Dashiell, J. F. *Fundamentals of general psychology.* Boston: Houghton Mifflin, 1937.
7. Dollard, J., & Miller, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.
8. Gibson, Eleanor J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.
9. Guthrie, E. R. *The psychology of learning.* (Rev. Ed.) New York: Harper, 1952.
10. Heidbreder, Edna. The attainment of concepts: I. Terminology and methodology. *J. gen. Psychol.*, 1946, **35**, 173–189.
11. Hollingworth, H. L. General laws of redintegration. *J. gen. Psychol.*, 1928, **1**, 79–90.
12. Hovland, C. I., & Kurtz, K. H. Experimental studies in rote-learning theory: X. Pre-learning syllable familiarization and the length-difficulty relationship. *J. exp. Psychol.*, 1952, **44**, 31–39.
13. Hull, C. L. Quantitative aspects of the evolution of concepts. *Psychol. Monogr.*, 1920, **28**, No. 1 (Whole No. 123).
14. Hull, C. L. The concept of the habit-family hierarchy and maze learning. *Psychol. Rev.*, 1934, **41**, 33–52.
15. Hull, C. L. *Principles of behavior.* New York: Appleton-Century-Crofts, 1943.
16. Kline, L. W. An experimental study of associative inhibition. *J. exp. Psychol.*, 1921, **4**, 270–299.
17. Lewis, D., McAllister, Dorothy E., & Adams, J. A. Facilitation and interference in performance on the modified Mashburn apparatus: I. The effects of varying the amount of original learning. *J. exp. Psychol.*, 1951, **41**, 247–260.
18. McGeoch, J. A. *The psychology of human learning.* New York: Longmans, Green, 1942.
19. Mandler, G. Transfer of training as a function of degree of response overlearning. *J. exp. Psychol.*, 1954, **47**, in press.
20. Morgan, R. L., & Underwood, B. J. Proactive inhibition as a function of response similarity. *J. exp. Psychol.*, 1950, **40**, 592–604.
21. Noble, C. E. An analysis of meaning. *Psychol. Rev.*, 1952, **59**, 421–430.
22. Osgood, C. E. The similarity paradox in human learning: a resolution. *Psychol. Rev.*, 1949, **56**, 132–143.
23. Porter, L. W., & Duncan, C. P. Negative transfer in verbal learning. *J. exp. Psychol.*, 1953, **46**, 61–64.
24. Rossman, Irma L., & Goss, A. E. The acquired distinctiveness of cues: the role of discriminative verbal responses in facilitating the acquisition of discriminative motor responses. *J. exp. Psychol.*, 1951, **42**, 173–182.
25. Sheffield, F. D. The role of meaningfulness of stimulus and response in verbal learning. Unpublished doctor's dissertation, Yale Univer., 1946.
26. Siipola, Elsa M., & Israel, H. E. Habit interference as dependent upon stage of training. *Amer. J. Psychol.*, 1933, **45**, 205–227.
27. Sperry, R. W. Neurology and the mind-brain problem. *Amer. Scientist*, 1952, **40**, 291–312.
28. Underwood, B. J. *Experimental psychology.* New York: Appleton-Century-Crofts, 1949.
29. Underwood, B. J. Proactive inhibition as a function of time and degree of prior learning. *J. exp. Psychol.*, 1949, **39**, 24–34.
30. Wylie, H. H. Transfer of response in the white rat. *Behav. Monogr.*, 1909, **3**, No. 5.

(Received August 10, 1953)

Psychological Review
Vol. 61, No. 4, 1954

KNOWLEDGE AND STIMULUS–RESPONSE PSYCHOLOGY[1]

D. E. BERLYNE

University of Aberdeen, Scotland

A frequent source of uneasiness among psychologists is the increasing recklessness with which the stimulus-response type of theory, after years of servitude in animal laboratories, is being let loose among some of human psychology's most cherished preserves (11, 22, 37, 42, 48, 49). Many have long felt (46, 53) that even what lower animals do depends in some sense on what they "realize" about their environment, so that a psychology which does not have cognitions or perceptions as its basic concepts is poorly equipped for the study of infrahuman species. And many more can see, with a prophetic confidence paralleled only among the early opponents of experimental psychology, that S-R theory cannot progress very far with human behavior, because human beings do not just perform blind, automatic reflexes; they know what they are doing, and their actions are guided by what they know.

However, the only way to find out whether a theoretical approach is doomed to failure or is premature is to try it out. If we can find a way to analyze the role of knowledge and related phenomena in S-R terms, we may derive several important benefits. We shall be able both to take advantage of the rigor and precision of S-R language and to bring into view the relations between higher mental processes and fundamental principles of mammalian behavior. It has been abundantly demonstrated that S-R terminology need in no way do violence to the flexibility and rationality of human activity. Accounts of reasoning have been offered (11, 26, 42) which do not appear to contradict radically either the facts adduced by those who have studied insightful problem solution or the tenets of S-R reinforcement theory. Discussion in similar terms of language processes, from which most of the psychological uniquenesses of human beings are generally held to spring, has been far from abortive (39, 49). And belated attempts to fit perception into the same framework have not encountered any immediate obstacle (3, 50, 52). The present paper discusses whether or not the human capacity for acquiring and using knowledge need escape the omnivorous maw of S-R behavior theory.

Since the concepts and principles which we shall apply to our topic originated in the investigation of very different problems, mostly in animal psychology, we shall have to rely on a procedure, which we shall call *concept extension*, that has often provoked misgiving. We shall propose that terms which were introduced to describe one class of phenomena—stimulus, response, drive, etc.—be extended to new classes of phenomena. Now it is frequently noted that human beings, and not least psychologists, are unduly prone to think that they have explained something when they have attached a new name to it. And cases of concept extension are apt to elicit the knowing look and the triumphant pounce from those who are commendably on the watch for such aberrations. But their well-meant vigi-

[1] This article arose out of work performed under the auspices of the Yale Communication Research Program, which is directed by Dr. C. I. Hovland and financed by a grant from the Rockefeller Foundation. It was written while the author was a member of the staff of Brooklyn College.

lance is here completely out of place, since concept extension is much more than mere labeling. The concepts which figure in systems such as Hull's behavior theory (29, 30) are defined by sets of relations with other variables. Applying them to new phenomena means therefore postulating that the same relations hold for these cases and thus laying down a set of hypotheses which can be followed until they prove inadequate. Concept extension is closely comparable to what happens in a court of law when it is ruled that taxis are hackney carriages or that gramophone records are a form of writing. This, far from being an idle verbal eccentricity, immediately applies a large body of traffic regulations or of law of libel to a large class of new instances, until further experience makes it desirable to pass new legislation.

"Knowledge" as a Concept in Behavior Theory

1. Ryle (47) has indicated that knowledge is a *dispositional concept*. An individual is said to possess it even when he is not in any way manifesting it, so that, like such terms as brittleness or electrical resistance, it expresses a probability that certain observable events ("truth-conditions" [10]) will occur, given certain additional conditions ("test-conditions" [10]). Dispositional concepts must, in a psychology which uses quantitative language, be represented by *intervening variables*, which are defined by describing the equations linking them to observable antecedent and consequent variables (29). At the present stage, our formulations cannot be too exact, since concepts only approach precise definition as a science progresses (10), but on the antecedent side, we can note that knowledge is a product of learning. Its strength depends on exposure to a stimulus situation in the past, on the performance of responses in that situation, and on certain motivating and reinforcing conditions, which have been discussed elsewhere (4, 5). We assume that the possibility of innate knowledge, which at one time interested rationalist philosophers, can be discounted. On the consequent side, we can bear in mind Skinner's point that "to know is largely to be able to talk about" (49). It is true that testing ability to produce verbal behavior is one of the most convenient operations for measuring amount of knowledge, as in the traditional scholastic examination. But it is not the only one. There are times when it is dangerous to judge how much a person knows from how much he talks, and a dumb man may know more than a windbag. There appear to be, in fact, three principal effects of knowledge which can be a source of consequent variables: (*a*) performance of verbal responses, (*b*) production of new knowledge (as in reasoning) or evocation of other implicit responses (e.g., attitudes or thoughts), and (*c*) effects on overt behavior. As regards the last-named, all we can say at this stage is that knowledge causes the organism to behave in some respects as it would if the events or objects which are known were present. To use the terminology favored by cyberneticians and by Skinner, it causes present behavior to be "controlled" by absent or past events. With this point of departure, we can proceed to narrow down step by step the class of variables to which knowledge can be assigned.

2. The intervening variables that constitute knowledge are *habits*. The responses mediated by these habits may be overt, e.g., speaking or writing, or they may be implicit, e.g., thinking.

3. The responses mediated by these habits are *cue-producing responses* (11) and, if implicit, they are furthermore

pure stimulus acts (23). This means that they produce self-stimulation which can influence subsequent behavior of the same organism. Subjects who "know about" particular events or objects are said to "be conscious of" them. It is worth noting that, although consciousness no longer enjoys the consideration that was held its due in the days of introspective psychology, it is still found necessary to distinguish conscious from unconscious processes, because of the special properties of overt behavior dependent on conscious processes. Behaviorist writers (23, 35, 38) have attributed such behavior, regarded as largely a human prerogative, to the capacity to react to one's own reactions. Those (e.g., 11, 17) who have been inclined to follow Freud's assertion (14) that the unconscious is unverbalized have particularly stressed reaction to self-stimulation from verbal responses. The intervention of cue-producing responses has, moreover, been the key admitting insightful problem solution to the domain of S-R theory (11, 26, 42).

We have at our disposal two definitions of a "symbol" in behavioral terms, one offered by Morris (39) and one by Osgood (43). For the former, a symbol is a *"sign that is produced by its interpreter and acts as a substitute for some other sign with which it is synonymous,"* and *"if anything, A, is a preparatory-stimulus which in the absence of stimulus-objects initiating response-sequences of a certain behavior-family causes a disposition in some organism to respond under certain conditions by response-sequences of this behavior-family, then A is a sign"* (39, p. 10). This definition does not seem altogether satisfactory, as it would make signs of drive-stimuli, drive-producing stimuli and anything that induces any sort of set. Osgood's definition obviates this and other objections: *"a pattern of stimulation which is not the object is a sign of the object if it evokes in an organism a mediating reaction, this* (a) *being some fractional part of the total behavior elicited by the object and* (b) *producing distinctive self-stimulation that mediates responses which would not occur without the previous association of nonobject and object patterns of stimulation"* (43, p. 204).

Osgood's scheme thus follows this pattern:

$$\boxed{S} \longrightarrow r_m - - - \rightarrow s_m \longrightarrow R_X,$$

where $\boxed{S}$ is the sign, r_m the mediating reaction (a part of the response pattern made to the signified object), and R_X is the overt behavior evoked by r_m. It will be seen that r_m is the behavioral equivalent of the "concept" or "meaning" which a sign, in traditional accounts of symbolization, "conjures up" in the interpreter's "mind." Its introduction is necessary for one very good reason, apart from those put forward by Osgood. Most studies of symbols in animals and men have naturally begun with the single symbol as a unit. But in human higher mental processes, especially in knowledge, the units are actually complex combinations or sequences of elemental symbols, e.g., sentences (propositions) or complex perceptual responses. Since these combinations or complexes have effects that their constituents would not have alone, we have an instance of *patterning* as described by Humphrey (32) and analyzed further by Hull (28, 29). But this principle by itself will not suffice. The sentence "man bites dog" will have a different effect from "dog bites man" but a similar one to "human being sinks teeth into canine animal." Yet, one might expect the opposite to be true, since the first two sentences must be nearer together on any primary generalization continuum. Similarly, on the response side, the same remembered material is likely to be expressed in different words on different occasions (2). So, different complexes of signs can come, through learning, to evoke the same "meaning" (which Osgood identifies by r_m), and the same meaning can come, through learning, to evoke different overt responses. We have therefore clear cases of "secondary stimulus generalization" (29) or "acquired equivalence of cues" (11) on the one hand and secondary response generalization, acquired equivalence of responses, or habit-family hierarchy (25, 30) on the other. These processes require a mediating cue-producing response (11, 25, 29).

When Osgood's r_m is evoked by a stimulus other than those coming from either the signified object or an external sign ("signal" [39]), whether such a stimulus be external or response-produced (e.g., by another r_m), we have what we shall hereafter call a symbol or symbolic response.

4. Knowledge mediates *believed* symbols. Morris (39) attempts a behavioral account of belief (which, he points out, can occur in differing degrees, as in judgments of probability) as the degree to which the organism is disposed to respond as if the signified object existed; Skinner (49) likewise defines belief in terms of strength of response. However, much of the response pattern conditioned to the significatum can occur even when the sign or symbol is disbelieved; it may very well evoke the same material in free association or directed thinking as it would if believed, and, as works of art and literature show, it may arouse similar emotional responses, though probably less intensely. It is rather in the *overt* behavior that we must look for a measure of belief, and it is principally this that is inhibited in doubt or disbelief.

Philosophers have, of course, long wrestled with the problem of distinguishing belief from knowledge. For some, knowledge is true belief, while others insist that, in addition to being true, beliefs must be supported by adequate evidence to justify the name of knowledge. These questions, though important and no doubt capable of empirical formulation, need not concern us here. False beliefs affect behavior just as they would if they were true, at least until something arises to make the subject doubt them, and we can presume that similar motivation underlies the absorption of knowledge and error. We shall therefore not draw a distinction between knowledge and belief.

5. The believed symbols mediated by knowledge are *designative*. Morris (39) classifies symbols according to the distinguishing characteristics of the objects or events which they signify. If these characteristics are stimulus properties, the symbol is *designative*, if they consist of a preferential status with respect to the organism's needs, it is *appraisive*, and if they take the form of a tendency to evoke certain sorts of overt behavior, it is *prescriptive*. We can translate this valuable classification into Osgood's scheme by categorizing symbols according to which fractional component of the significatum's response pattern has come to form the r_m. If the r_m is composed mainly of emotional and drive-producing responses, it is an appraisor, whereas if it consists largely of fractional skeletal responses, it is a prescriptor. Designators will thus be those symbols which are built up of responses dependent on the stimulus properties of the significatum, and these may consist of perceptual responses (3) or verbal responses of the kind Skinner calls "tacts" (49).

Trains of Thought

Many writers have described how the human being's responses depend jointly on the present stimulus situation and, with recent experience predominating, on a whole mass of relevant past experience which has left traces in his nervous system. We have Herbart's apperceptive mass (21), Bartlett's schema (2), and the modern social and perceptual psychologist's frame of reference (20) as concepts referring to this phenomenon. Moreover, it has been pointed out that when these traces, which underlie knowledge, are reactivated, they give rise to long sequences of intraorganismic events —Bartlett's schema (2), Hull's sequences of pure stimulus acts (23), and Hebb's phase sequences (19).

There is no reason why we should not give to these sequences, which can bring about both the recall of old knowledge and, as in reasoning, the production of new knowledge, their everyday name, *trains of thought.* We can, by drawing on Hull (23, 29) and Bartlett (2), suggest the following six stages by which trains of thought may have developed out of simple response capacities in an animal as well equipped with ability to symbolize as the human being:

1. *Reaction.* Hull starts his account (23) by describing a series of events in the external world, which produce a parallel series of events or reactions in an organism. But, as he acknowledged in a footnote (p. 512), this account is deficient. The organism's reactions depend not only on external events but also on certain intervening variables representing conditions inside the organism; chief among these are the effects of previous learning ($_{S}H_{R}$) and motives (D). This explains why different individuals not only perform different overt responses to the same external stimuli, but also derive different perceptions and later different knowledge from exposure to identical situations.

2. *Redintegration.* If S_1, S_2, etc. habitually occur in the same order or simultaneously, then foresight or expectancy can emerge. Both S_1 and s_1 (the proprioceptive stimulus resulting from R_1) can become conditioned to at least a fractional component of R_2. These components can include incipient skeletal responses (postural sets) and visceral responses such as fear (36, 41), but also, and these are what concern us most, perceptual and subvocal cue-producing responses.

If S_1 has come in this way to evoke, fractionally or subliminally, the perceptual response ($\bar{r}_2$) appropriate to S_2 (3), then some familiar phenomena from the psychology of perception follow from the principles of behavior theory.

a. If the habitual S_2 follows or accompanies S_1, then the principle of summation of reaction potentials (29, Corollary v) leads us to expect a lowering of the threshold for perceiving S_2. There will be values of the reaction potentials $_{S_1}E_{\bar{r}_2}$ and $_{S_2}E_{\bar{r}_2}$ such that, although neither exceeds the reaction threshold ($_{S}L_{R}$) separately, the behavioral sum of the two will. In any case, the sum of both will be greater than $_{S_2}E_{\bar{r}_2}$ alone. Thus, this increase in perceptual reaction potential, which the writer has elsewhere identified with attention (3), explains the familiar fact that expected events are more likely to be perceived than others and are likely to be perceived more vividly. It is the phenomenon which Hebb calls the "central reinforcement of a sensory process" (19).

b. If S_2 is an ambiguous stimulus, i.e., if it evokes two or more incompatible perceptual response tendencies of about equal strength, then the one reinforced by the expectancy conditioned to S_1 will prevail. This is one case of the influence of set on perception (9, 33, 52).

c. If the habitual S_2 is for once replaced by a somewhat different stimulus, S_A, then several results might ensue:

(i) *Illusion* or *dominance* (8, 9, 45). S_A will evoke $\bar{r}_2$ by stimulus generalization, response generalization, or both, although this will normally be less strong than $\bar{r}_A$, the "accurate" perception (i.e., the most frequent perceptual response to S_A). But $\bar{r}_2$, when strengthened by the expectancy aroused by S_1, may well prevail, so that S_A will be wrongly perceived as if it were S_2.

(ii) *Compromise* (8, 9). Both S_1 and S_A may evoke, by generalization, a perceptual response tendency corresponding to some stimulus occupying an intermediate position between S_A and S_2 on a continuum. In that case the strength of this compromise perception, contributed to by both, may exceed that of the expected perception ($\bar{r}_2$) and that of the accurate one ($\bar{r}_A$).

(iii) *Raised threshold.* If $\bar{r}_2$ is not strong enough to prevail over $\bar{r}_A$, it may interfere with it in such a way as to reduce its effective reaction potential (${}_{s_A}\bar{E}_{\bar{r}_A}$) (8, 45) and make its perception less probable.

d. If the habitual S_2 is absent, and the S_A which replaces it is so remote from it that no compromise or illusion is possible (i.e., because the generalized ${}_{s_A}\bar{E}_{\bar{r}_2}$ is too weak), then two cases can arise:

(i) In conditions of poor visibility where S_A cannot be seen clearly (i.e., where ${}_{s_A}E_{\bar{r}_A}$ is weak), the $\bar{r}_2$ conditioned to S_1 may be supraliminal by itself. In that case we shall have a *hallucination* of S_2. This happens when tachistoscopic figures are falsely completed (2) and when hallucinations are produced by suggestion (40) or conditioning (12).

(ii) In conditions of good visibility, where the discrepancy between S_2 and S_A cannot be overlooked, we shall have *conflict* between the incompatible perceptual responses, $\bar{r}_A$ of peripheral origin and $\bar{r}_2$ of central origin. If we assume (7) that conflict is a drive condition (C_D), this explains the emotional effect that results from the clash between an expectancy and an external stimulus and plays a great part in Hebb's theory (18, 19).

3. *Symbolization.* Our consideration of stage 2 reveals the role played by previous knowledge in perception.[2] But a step forward is achieved when two new conditions are fulfilled: (*a*) s_1, the proprioceptive stimulus produced by R_1, is sufficient without S_1 to evoke a fractional component of R_2; and (*b*) this fractional component can be supraliminal without support from S_2. Then the fractional component of R_2 can become a symbol (r_m) for S_2 and represent it in its absence. We then have the possibility of a true train of thought, a sequence of internal responses (symbols) which can act in lieu of a remembered, anticipated, or imaginary series of external events. Each symbol is in its turn elicited by the response-produced cue of the previous symbol, so that a behavior chain (30) is formed, comparable to those of temporal maze habits (51, 54) or human rote memory (27, 31). The symbols (r_m) constituting such trains of thought may include perceptual responses ($\bar{r}$) or subvocal verbal responses (r_v).

4. *Ramification.* The next complications arise when S_1 participates in several habitual sequences of events at different times and thus can initiate several alternative associated responses. Similarly, each symbol in its turn may be able to lead the train of thought off in many alternative directions. But what determines precisely which response out of the many alternatives occurs? From Hull's account (30, p. 312) we can expect four factors to determine it jointly:

[2] It should be pointed out that, just as knowledge acquisition is best regarded as one special sort of learning, namely that sort which enables a symbolic response to act in lieu of an absent stimulus, so the role of knowledge in perception does not exhaust the role of learning in perception. The cases we have been considering are those where the perception of a stimulus is supplemented or replaced by components of its perceptual responses which have been conditioned to cues habitually accompanying it. There appear to be many other cases where learning affects perception quite differently: a stimulus which could give rise to a number of alternative perceptions gives rise to one in particular because that one has been reinforced more than the others (34, 52).

a. External stimuli (S): In the case of autonomous trains of thought, only one of these is required in order to initiate the sequence. We shall therefore refer to such starting points as *initiating stimuli.*

b. Response-produced stimuli (s). These may be proprioceptive cues from muscular responses (s_p), or response-produced cues from perceptual responses ($\bar{s}$) or verbal responses (s_V). They keep the train of thought going after the initiating stimulus has ceased.

c. Drive-stimuli (S_D). These continue throughout the sequence until the drive has been reduced and so become conditioned to every response in the chain. But the reinforcement-gradient principle implies that they will be most strongly conditioned to responses coming just before reinforcement.

d. Fractional goal-stimuli (s_G). These are internal cues produced by fractional anticipatory goal responses (r_G). They have the dual function of providing secondary reinforcement for earlier responses in the series (**30**, Corollary xv) and directing the series toward a goal (**24**).

The first two of these factors we shall call *cue-stimuli* and the last two we shall call *motivational stimuli,* noting that the two pairs have somewhat different roles. The cue stimuli provide the starting point for a train of thought and restrict the future course of the sequence to the relatively narrow range of responses to which they are conditioned. The motivational stimuli are conditioned to a much wider range of responses, since they must have coincided with an enormous variety of situations; they accordingly select from the repertoire made available by the cue-stimuli those items which are likely to contribute most effectively to the satisfaction of the motives aroused, and, in general, they serve to keep the train of thought on a path leading to the solution of the problem on hand. In addition, the drives with which they are associated impel the chain of symbols to continue until the drives have been reduced or extinction has supervened.

The above conception has been derived from studies of maze learning in rats. It is therefore encouraging to note that other writers have been driven to recognize two corresponding sets of factors, as a result of direct attacks on higher mental processes in human beings.

Why only the correct association appears, whether it be a question of a single reaction, as in the controlled-association experiment, or of long successions of thoughts, as in directed thinking, was one of the principal interests of the Würzburg school. Their pursuit of the answer culminated in the theory put forward by Ach (1). It depends, he said, on the presence in consciousness of an "idea of the stimulus" (*Reizvorstellung*) and of an "idea of the aim" (*Zielvorstellung*). It is not hard to see in these two concepts the impact on the organism of cue-stimuli and motivational stimuli, respectively. They jointly produce a "determining tendency," which acts to steer the thought sequence toward the aim and to exclude irrelevant digressions.

Again, in Bartlett's theory of remembering, recall is the product of both the stimulus which elicits the remembering process (which "reminds" one) and what he calls an "attitude," which he describes as "very largely a matter of feeling or affect" (2). The latter ensures that the material which emerges is something pertinent to the present situation and not just a fortuitous association. Our cue-stimuli and motivational stimuli have thus obtruded themselves in yet another guise.

5. *Reorganization.* An important advance is accomplished when the symbols making up trains of thought are no longer tied to one chronological order but become capable of rearrangement.

This added flexibility makes possible "the assembly of behavior segments in novel combinations suitable for problem solution" (**26**; **30**, ch. 10), and thinking can perform "the two different functions of preparation for reality (anticipation of what is probable) and substitution for reality (anticipation of what is desirable)" (**13**, p. 50).

The process of "short-circuiting" or "serial-segment elimination" presupposes, according to Hull (**25**), some persistent stimulus which acts during the whole of the sequence. Such a stimulus can become more strongly conditioned to later than to earlier items in the sequence, by virtue of the reinforcement gradient, and can thus serve to elicit anticipatorily those responses which immediately precede reinforcement, so that they crowd out irrelevant and unhelpful diversions. Internal events, and especially those we have termed motivational stimuli, serve this purpose.

Closely related conceptions appear in the writings of Hebb (**19**) and Bartlett (**2**). The former describes how the evocation of a familiar and long-established phase sequence comes in time to mean simply a review of its highlights, the less important connecting material gradually dropping out. Bartlett attaches an extreme importance to the ability of human organisms to "turn round on their own schemata," i.e., to "go directly to that portion of the organized setting of past responses which is most relevant to the needs of the moment" (**2**, p. 206). This ability obviates the necessity of reviewing a succession of trivial memories in order to reach the point of time which is important, as happens in some primitive forms of remembering. The factors responsible are "interest, appetite, etc." These are obviously motivational terms, and once again we can see an instance of motivational stimuli leading straight to those responses which are most "relevant" to them (i.e., most closely contiguous with their cessation). Bartlett also describes the formation of specialized "schemata" (or organized system of retained material) pertaining to particular "appetites, instinctive tendencies, interests and ideals." Thus, once more we find attributed to motivational stimuli the power to tie together, and thus make readily available in close succession, those response tendencies which are most likely to subserve particular drives or purposes.

6. *Ratiocination.* The final refinement in the human being's application of knowledge is logical or, as Piaget (**44**) calls it, "operational" thinking. For this, the organism has to learn to perform only such symbol sequences as fulfil certain conditions ("rules of logic") which are necessary to ensure their stability and consistency. Some of these conditions are enumerated by Piaget, who outlines the stages by which a child gradually comes to achieve them. The reinforcement for this learning seems to come both from social reward and from the better adapted (more "intelligent") behavior that logical thought makes possible. Except when the restrictions of realistic reasoning are suspended—as in dreams and fantasy (**16**), wit (**15**), etc.—fortuitous, irrelevant, or illogical associations are inhibited. This is presumably because stimuli produced by such responses evoke some sort of acquired drive (e.g., Dollard and Miller's "learned drive to make . . . explanations and plans seem logical" [11, p. 120]).

If, as is hoped, this discussion violates neither the nature and importance of knowledge nor the findings of S-R learning theory, we can use the latter as a valuable source of hypotheses with which to attack many central problems in the higher mental processes. As an example, this account has given rise to a theory and to some experimental work on the much neglected topic of human curiosity, the motivation behind the acquisition of knowledge (**4, 5, 6**).

Summary

An attempt is made to conceptualize knowledge in stimulus-response language. Knowledge, according to this analysis, consists of habits which mediate believed, designative symbols. It is suggested that symbol sequences or trains of thought are likely to have developed through six stages from the simplest response capacities to logical thought. Some of the phenomena that are familiar to investigators of thinking and perception are shown to be consonant with this account.

REFERENCES

1. Ach, N. *Über die Willenstätigkeit und das Denken.* Göttingen: Vandenhoeck & Ruprecht, 1905.
2. Bartlett, F. C. *Remembering.* Cambridge: Cambridge Univer. Press, 1932.
3. Berlyne, D. E. Attention, perception and behavior theory. *Psychol. Rev.*, 1951, **58**, 137–146.
4. Berlyne, D. E. Some aspects of human curiosity. Unpublished Ph.D. thesis, Yale Univer., 1953.
5. Berlyne, D. E. A theory of human curiosity. *Brit. J. Psychol.*, in press.
6. Berlyne, D. E. An experimental study of human curiosity and its relation to incidental learning. *Brit. J. Psychol.*, in press.
7. Brown, J. S., & Farber, I. E. Emotions conceptualized as intervening variables—with suggestions toward a theory of frustration. *Psychol. Bull.*, 1951, **48**, 465–495.
8. Bruner, J. S., & Postman, L. On the perception of incongruity: a paradigm. *J. Pers.*, 1949, **18**, 206–223.
9. Bruner, J. S., Postman, L., & Rodrigues, J. Expectation and the perception of color. *Amer. J. Psychol.*, 1951, **64**, 216–227.
10. Carnap, R. Testability and meaning. *Phil. Sci.*, 1936, **3**, 420–471; 1937, **4**, 1–40.
11. Dollard, J., & Miller, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.
12. Ellson, D. G. Hallucinations produced by sensory conditioning. *J. exp. Psychol.*, 1941, **28**, 1–20.
13. Fenichel, O. *The psychoanalytic theory of neurosis.* New York: Norton, 1945.
14. Freud, S. The unconscious. In *Collected papers.* Vol. IV. London: Hogarth, 1925. Pp. 98–136.
15. Freud, S. *Wit and its relation to the unconscious.* In A. A. Brill (Ed.), *The basic writings of Sigmund Freud.* New York: Modern Library, 1938. Pp. 633–803.
16. Freud, S. *A general introduction to psychoanalysis.* New York: Permabooks, 1953.
17. Guthrie, E. R. *The psychology of learning.* New York: Harper, 1935.
18. Hebb, D. O. On the nature of fear. *Psychol. Rev.*, 1946, **53**, 259–276.
19. Hebb, D. O. *The organization of behavior.* New York: Wiley, 1949.
20. Helson, H. Adaptation-level as a basis for a quantitative theory of frames of reference. *Psychol. Rev.*, 1948, **55**, 297–313.
21. Herbart, J. F. *Psychologie als Wissenschaft, neu gegründet auf Erfahrung, Metaphysik und Mathematik.* Königsberg: Unzer, 1824–1825.
22. Hovland, C. I., Janis, I. L., & Kelley, H. H. *Communication and persuasion.* New Haven: Yale Univer. Press, 1953.
23. Hull, C. L. Knowledge and purpose as habit mechanisms. *Psychol. Rev.*, 1930, **37**, 511–525.
24. Hull, C. L. Goal attraction and directing ideas conceived as habit phenomena. *Psychol. Rev.*, 1931, **38**, 487–506.
25. Hull, C. L. The concept of the habit-family hierarchy and maze learning. *Psychol. Rev.*, 1934, **41**, 33–52; 134–152.
26. Hull, C. L. The mechanism of the assembly of behavior segments in novel combinations suitable for problem solution. *Psychol. Rev.*, 1935, **42**, 219–245.
27. Hull, C. L. The conflicting psychologies of learning—a way out. *Psychol. Rev.*, 1935, **42**, 491–516.
28. Hull, C. L. Words and their contexts as stimulus aggregates in action evocation. Unpublished memorandum, Yale Univer. Medical Library, 1941.
29. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
30. Hull, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.
31. Hull, C. L., Hovland, C. I., Ross, R. T., Hall, M., Perkins, D. T., & Fitch, F. B. *Mathematico-deductive theory of rote learning.* New Haven: Yale Univer. Press, 1940.
32. Humphrey, G. *The nature of learning.* New York: Harcourt, Brace, 1933.

33. Kilpatrick, F. P. (Ed.) *Human behavior from the transactional point of view.* Hanover, N. H.: Institute for Associated Research, 1952.

34. Kohler, I. Über Aufbau und Wandlungen der Wahrnehmungswelt. *Österr. Akad. d. Wiss., Phil-hist. Klasse, Sitzungsber. 227, 1 Abh.,* 1951.

35. Lashley, K. S. The behaviorist interpretation of consciousness. *Psychol. Rev.,* 1923, 30, 237–272; 329–383.

36. Miller, N. E. Studies of fear as an acquirable drive: I. Fear as motivation and fear-reduction as reinforcement in the learning of new responses. *J. exp. Psychol.,* 1948, 38, 89–101.

37. Miller, N. E., & Dollard, J. *Social learning and imitation.* New Haven: Yale Univer. Press, 1941.

38. Morris, C. W. Foundations of the theory of signs. *Int. Encyc. unif. Sci.,* 1938, 1, No. 2.

39. Morris, C. W. *Signs, language and behavior.* New York: Prentice-Hall, 1946.

40. Mowrer, O. H. Preparatory set (expectancy)—a determinant in motivation and learning. *Psychol. Rev.,* 1938, 45, 62–91.

41. Mowrer, O. H. A stimulus-response analysis of anxiety and its role as a reinforcing agent. *Psychol. Rev.,* 1939, 46, 553–565.

42. Mowrer, O. H. *Learning theory and personality dynamics.* New York: Ronald, 1950.

43. Osgood, C. E. The nature and measurement of meaning. *Psychol. Bull.,* 1952, 49, 197–237.

44. Piaget, J. *La psychologie de l'intelligence.* Paris: Colin, 1947. (*The psychology of intelligence.* New York: Harcourt, Brace, 1950.)

45. Postman, L., Bruner, J. S., & Walk, R. D. The perception of error. *Brit. J. Psychol.,* 1951, 42, 1–10.

46. Ritchie, B. F. The circumnavigation of cognition. *Psychol. Rev.,* 1953, 60, 216–221.

47. Ryle, G. *The concept of mind.* New York: Barnes & Noble, 1949.

48. Skinner, B. F. *Science and human behavior.* New York: Macmillan, 1953.

49. Skinner, B. F. *Verbal behavior.* (William James Lectures, Harvard University, 1947.) Cambridge: Harvard Univer. Press, in press.

50. Spence, K. W. Cognitive versus stimulus-response theories of learning. *Psychol. Rev.,* 1950, 57, 159–172.

51. Spragg, S. D. S. Anticipatory responses in serial learning by chimpanzee. *Comp. Psychol. Monogr.,* 1936, 13, No. 62.

52. Taylor, J. G. *The behavioural basis of perception.* In press.

53. Tolman, E. C. *Purposive behavior in animals and men.* New York: Century, 1932.

54. Woodbury, C. B. Double, triple and quadruple repetition in the white rat. *J. comp. physiol. Psychol.,* 1950, 43, 490–502.

(Received August 13, 1953)

Psychological Review
Vol. 61, No. 4, 1954

THE CONCEPT OF INTELLIGENCE AND THE PHILOSOPHY OF SCIENCE

CHARLES C. SPIKER AND BOYD R. McCANDLESS

Iowa Child Welfare Research Station

A careful application of the principles of the philosophy of science to controversial issues within an area of an empirical science has often proved clarifying. These methodological (logical) analyses have occasionally demonstrated that some of the questions which scientists considered appropriate for experimental attack could actually be resolved only after linguistic analysis. The major contribution of any such analysis is the reformulation of some of the traditional questions. The present paper attempts such an analysis of the psychological concept, "intelligence."

The paper is presented in two parts. The first contains a summary of the important points, relevant to this analysis, of the frame of reference within which the writers evaluate the methodological problems of their science. Writers of the philosophical school of "logical positivism," or "scientific empiricism," have written explicitly on the methodology of psychology, formulating principles that may be regarded as the fundamental principles of neo-behaviorism (2, 3, 5, 8). The second part deals with the application of these philosophical principles to problems associated with the investigation of human intelligence.

The Methodological Frame of Reference

The principles that scientists have followed in the formulation of their concepts have been made explicit by philosophers as a result of their analyses of the language of science, of which the language of the physical sciences is the prototype. The language of science is a physicalistic language; that is, the referents of the descriptive terms occurring in scientific discourse are physical objects or events, their properties, and their relationships. There is, therefore, implicit in the philosophy of scientists a basic assumption regarding a "real world." The scientist assumes that there is a blueness "out there" when he has a sensation of blue. This "naive realism" of the scientist is not to be confused with any metaphysical viewpoints with reference to the nature of "reality." The scientist's position in this respect may be regarded as a convenient working assumption. It is simply another way of stating his belief that the data with which he deals have sufficient generality and significance to warrant further study.

Concepts that have been accepted in science and have proved useful for theoretical reasons, and for more pragmatic reasons as well, can be defined so that they are reducible to very simple terms, which have been designated by Carnap as *primitive predicates* (6). This class of terms is distinguished in part by the fact that they cannot be further reduced, in the sense that they cannot be given *linguistic* definitions; understanding of such terms can be obtained only through acquaintance with their referents. While philosophers have not troubled to delimit this class of terms categorically, its important characteristics may be given by a few examples. There are the property or quality terms such as "blue," "green," "bright," "hard," etc.; the relational terms such as "to the left of," "above," "between," "brighter than," etc.; and, of course, a subclass of terms naming physical objects and events.

We may point out, parenthetically, that in scientific practice, concepts are not ordinarily reduced to (defined in terms of) such simple concepts. This would be laborious and, except for certain formal purposes, unprofitable. Words that may be reduced relatively easily to such a level are used without explicit definitions. Let us use the term "abstract" to refer to words whose definitional chains are long in the sense that numerous statements are required for defining them solely in terms of the primitive predicates. We may then describe scientific *practice* in this regard as that which utilizes explicit definitions only for the more abstract concepts; (even in these cases, the reduction process is carried down only so far as is necessary to avoid serious ambiguity). Such a statement, and rightly so, does not specify a crucial or necessary length of the definitional chain in order that the concept thereby defined be an abstract one.

If each acceptable term in a scientific language can be defined with reference to such terms as "blue," "above," "hard," etc., then the concepts in science refer in the last analysis to things that are *immediately observable* in a very simple sense of this italicized phrase. It is just this characteristic of scientific language which is intended when it is said that the language of science is a physicalistic language or that it has a physicalistic verification basis.

The formation of scientific concepts may be best understood through an exposition of the grammatical (logical) form of definitions in general, technically known as "definitions in use." Conventionally, one finds on the left side of an equation-like arrangement of two sentences a sentence in which the term to be defined occurs. This sentence ordinarily states one of the simplest things that can be said about the term to be defined. For example, we may wish to define the concept "length." On the left side we may write the simple statement, "the length of this table is five feet." In the more important definitions, there is on the right-hand side of the definition a statement (or set of statements) that presents a relatively complex set of interrelationships among other terms, typically of the form: "If . . . then - - -." The two statements, the one on the right and the one on the left, are then connected by a symbol which carries the meaning, "means by verbal agreement the same as." If we fill in the right-hand side of the definition, the above statement about length means the same as: "*If* one takes a foot rule and repeatedly places it so that there is no gap and no overlapping of one placement and another, and if each placement is parallel to the edge of the table, *then* five such placements may be made between the edges perpendicular to the direction of the placements." The meaning of length is not explicitly carried, of course, unless the right-hand statement contains only terms which are already meaningful.

The groundwork has now been laid for an exposition of the phrase that has become so popular among psychologists —"operational definition." Bergmann (2) has pointed out that this term refers to nothing more complex than that science requires all terms occurring on the right-hand side of a definition to be, or to be reducible to, the primitive concepts we have already discussed. This requirement may be designated the *empiricist meaning criterion,* thereby avoiding some of the confusions which have become associated with "operationism." In order for a word to be meaningful by this criterion, it must be reducible, in the sense discussed above, to primitive predicates.

Obviously, this discussion describes an ideal procedure. One may look vainly through some introductory physics textbooks for an *explicit* definition of the concept of "mass." What one ordinarily finds are several statements about mass, any one of which might, according to our discussion of meaning, be construed as a definition. This fact points up the need for considering the second methodological principle concerning scientific concepts. Analyses of scientifically acceptable concepts show that these concepts not only meet the empiricist's meaning criterion, but in addition are lawfully related to other meaningful concepts—such relationships being exemplified by statements of the form: "If A, then B," where A and B are both meaningful concepts. In general, the more relationships a given concept has to other concepts, the more fruitful or useful it is said to be. Thus, in physics, the concepts of time, force, energy, mass, distance, etc. are extremely useful since they enter in some form into all laws of mechanics.

Many discussions of operationism have been found objectionable by some scientists—particularly by some psychologists—because they have not emphasized this second aspect of scientific concepts. The scientist may insist that his term "means" more than just what is contained on the right-hand side of any definition of it. To anticipate later discussions, he may insist that intelligence means more than just an IQ from a given test: a high "amount" for a given individual means that this individual will probably do well in school, is probably good at arithmetic, is not likely to be found in an institution for the feebleminded, probably has parents with high average school achievement, and the like. The present formulation does not rob the scientist of the richness of his "meaning." This *additional* meaning is carried by the statements of relationships between the unambiguously defined concept and the other concepts (i.e., school achievement, institutionalization, level of parental education, etc.). Conveniently, Bergmann (2) distinguishes between meaning I (formal, operational meaning) and meaning II (significance, usefulness, fruitfulness). A concept that does not meet the first criterion cannot meet the second. A concept that meets only the first criterion will eventually be discarded as useless.

Since scientists are usually not so formal and explicit as are philosophers about such matters, one frequently finds in a scientific discipline useful concepts for which formal definitions have not been given. In some such cases, it is possible to formulate two or more equally correct and equally simple definitions. The question of which definition to select for a given purpose is therefore a matter of convenience. It is not consistent, *in a formal sense,* to speak of alternative definitions for a concept, since an unambiguous term can have only a single definition within the same context; but one may speak loosely of a number of concepts in science for which, in practice, several definitions are possible. This fact merely points out that it often happens in science that two or more grammatically different definitions may define concepts which are so highly interrelated that it is convenient to give each set of referents the same name. In other words, the relationships between each of these formally different concepts on the one hand, and other concepts on the other, are, within acceptable limits of error, identical. It makes little difference for most purposes which concept is used. A case in point is the concept of electric current, which may be quantitatively defined in terms of the deflection of a magnetic needle, the amount of heat generated, or the amount of silver deposited in a solution of nitrate of silver. When such clearly invariant relationships are found, it is often tempting (and, perhaps, of heuristic value) to speak and think of the concept involved as if it referred to a "thing" ontologically independent of all the sets of operations, the description of any one of which could serve as the definition of the concept. It is usually implied in such discussions that the "thing" itself cannot be directly sensed, but that we "infer" its existence from the observable evidence (i.e., from the pattern of invariant relationships among the operationally defined concepts). Hence, it would be said, we may *measure* electricity, even though we cannot directly sense it, in much the same way that we might assemble evidence concerning the existence and size of a hidden room in a house by comparing external measurements of the building with measurements of the observable rooms in it. It should be apparent from what has been said previously that this is merely a manner of speaking, and like many metaphorical expressions, generates little confusion unless one begins to accept its literal meaning. In the latter event, scientifically sterile arguments arise as to what the "thing" would look like if we *could* directly sense it, or as to what the "correct" way is to measure (define) it.

It frequently happens in the development of a science that a word appearing in the everyday, common-sense language is taken into the language of that discipline and is given a new definition. In most such cases, the new meaning is in some sense similar to the meaning of the word in the ordinary language. The words "force" and "mass," for example, occurred in the English language before they were utilized in Newtonian physics. Most high school students of introductory physics learn to distinguish between the two meanings such words have, and little confusion seems to result. In the newer sciences, however, attempts are often made to convey factual information through the use of words from the ordinary language without explicit redefinition of such concepts. In extreme cases, it appears that some scientists, particularly those in the social sciences, conceive of science as a technique for "measuring" the things to which many of the words in the ordinary language presumably refer. While it is not the writers' intention to depreciate the usefulness of common-sense knowledge, they wish to point out that if it had no limitations, scientific knowledge would not be necessary. Also, if the language of common sense were sufficiently precise, it would be unnecessary to study mathematics and logic. In many cases it appears that attempts to quantify (redefine) words from the natural language are uneconomical. Many such concepts refer in a vague way to highly complex sets of interrelations among distinguishable phenomena. It appears that the most economical way to study such patterns would be to define several concepts referring to these phenomena, with subsequent attempts to make explicit by empirical investigations the interrelationships holding among them. An all too frequent substitute for such a procedure consists of an attempt to "capture" all the phenomena and relationships in the definition of a single concept.

The complaint is not infrequently heard that if one subscribes to operationism, he places severe and perhaps crippling limitations upon the extent of the generalizations he can make. The argument proceeds along the following lines: Suppose a psychologist does a series of experiments on the learning of a task under certain conditions, using adult human subjects, and concomitantly defines a concept that he calls "habit$_1$." Operationally, the definition of this term includes references to the specific task, the conditions of learning, and the human subjects. Now, if he changes the task and the conditions, he must, according to the principles of operationism, define a new concept,

"habit_2." If he keeps the same task and conditions, but uses chimpanzees, he must again define a new concept, "habit_3." Obviously—the argument continues—such a procedure requires an inconvenient number of terms. Thus, operationism is too stringent and places too many restrictions upon scientific generalization. Since the business of science is the discovery of general laws, operationism defeats the purpose of science.

There are two distinct issues involved in the preceding argument. First, no one would argue that the subscripts to the above concepts do not have discriminable referents, and phenomena which *can* be reliably discriminated *may*, if one's purpose requires it, be given different names. Scientific practice may not typically be so formal as to apply subscripts to the terms, but it does differentiate among habits as studied in T mazes, in Skinner boxes, or in classical conditioning situations. Therefore, second, the question actually is whether a differentiation among such referents, either by name or by description, is a convenience or a hindrance. Concept analysis may be useful in pointing to the gaps in factual information where more careless terminological usage has obscured this lack. While it may point out logical differences among several concepts, it cannot indicate when there is sufficient empirical evidence to collapse these several concepts into a single one, or, more precisely, when it is possible and useful to define a more general concept which incorporates subsidiary concepts previously defined. Much of what is called theory in present-day psychology represents attempts to formulate more and more general concepts, whether they be called "habit," "drive," "aggression," "sign-gestalt-expectations," or what. In this last respect, scientists, without aid from the methodologist, are generally on guard against what Bergmann calls "that spurious comprehensiveness which is paid for by vagueness and triviality" (3, p. 438).

A similar objection to operationism probably arises from a failure to understand the formal (analytic) approach utilized by many writers in the exposition of this principle. The logician instructs us that a definition is arbitrary in the sense that it is the designation of a symbol (word) as a representation of an idea or complex set of ideas; which particular symbol is selected is of no formal importance; what is important is that the relationship between the word and its meaning be made clear and explicit. There is no empirical connection between a word and its referent. Objections to this formulation often take a form that suggests some type of word fixation or "concretism." It seems doubtful that such a mode of thought actually underlies many of these objections. What such people probably intend to emphasize—and logicians would be the first to agree—is that, in science, concepts are defined for some purpose. The scientist always wishes to define his concept in such a way that it will have a factual exemplification; that is, the referent of the term must exist in the same way that the referent of "chair" exists. Moreover, the scientist wants his concept to enter into statements of laws—in many cases, to enter only into certain laws. These two requirements depend upon factual matters for their realization. Thus, when the logician says that definitions are purely arbitrary, he speaks from a formal point of view and does not intend anything so nonsensical as that empirical considerations do not enter into the scientist's selection of a particular definition. It should be apparent that the answers to this objection, as well as to the one just previously stated, constitute restatements of the Meaning I—Meaning II distinction in slightly different guises.

The reader may note in this section of the paper an omission of any discussion of measurement and quantification. Since intelligence testing has been traditionally associated with such matters ("mental measurement"), this omission may be regarded by some as serious. The writers offer three reasons for their decision: First, the over-all logic of measurement, especially in psychology, has been clearly set forth by Bergmann and Spence (4). Second, the internal logic of test construction, together with its most widely accepted methods and techniques, has been comprehensively covered in such articles as that by Bechtoldt (1) and others. Finally, the writers consider this problem unessential to the understanding of the broader logic of the concept of intelligence, the primary concern of this paper. Misconceptions concerning the additivity of IQ points, the equality of units, the normal distribution of intelligence, etc. probably do not frequently occur among workers who are well grounded in the logic of statistics and measurement, and much of the confusion may be expected to disappear with improvement in such training.

The Analysis of Intelligence

The term "intelligence" is one of a number of words that psychologists have taken from the natural language. Its common-sense meaning, like that of many similar concepts, is complex and indefinite. An unequivocal characterization of the common-sense notion is probably both impossible and unprofitable. Reflection on the common-sense meaning of intelligence, however, leads to the discovery of two important points: First, the meaning leads to logical contradiction since, on the one hand, an individual may be regarded as generally bright, and on the other, an individual may be considered intelligent with respect to one thing and unintelligent with respect to others. The second point is that the common-sense meaning of intelligence always refers to behavioral consistency. There is the implication that the behavior of the individual is in some way trans-situational. Intelligence, in the common-sense usage, is not a momentary state of the individual, but transcends to some degree the specific situations in which the individual behaves.

In reading the nonexperimental ("theoretical") literature concerning intelligence, one must conclude that much time and energy have been devoted to attempts to capture and make explicit the several connotations of the natural language concept. Such attempts have probably stimulated much research. It is the writers' opinion, however, that numerous sterile controversies and confusions have arisen from an inadequate analysis of the goals and purposes of work on intelligence.

The organization of intelligence. There is one important assumption common to all the frames of reference in which intelligence tests have been constructed, from Binet to the present day. This is the assumption of trans-situational consistency of behavior. However, the different emphases of different test constructors have drawn attention to the inconsistencies of the original common-sense notion of intelligence. Some have argued that there is a general intelligence, that the trans-situational consistency in the level of behavior extends to all situations requiring "intellectual" problem solving. The term "intellectual" has actually been defined by the items selected for the tests rather than by attempts to circumscribe the "population" of intellectual behavior. But others, utilizing factor analysis as a tool, see no a priori limitations to the number of factors required to account for the variability of "intellectual behavior" (e.g., Thurstone [10]). For them, the empirical data determine the number of factors. Still another group of investigators has consistently distinguished between "verbal" and "performance" intelligence, or between "abstract" and "concrete" intelligence.

It seems correct to state that no one, in any of these groups, has unambiguously circumscribed the population of "intellectual behavior" or has provided explicit sampling criteria for the selection of items for his tests. While this seriously limits the significance and objectivity of the frames of reference ("theories") in which the tests were said to be constructed, it does not detract in any way from any success in prediction that has been achieved by means of the tests; that is, the descriptions of the finished tests, and the accompanying instructions for administering and scoring them, constitute formally satisfactory definitions of the several concepts of intelligence, despite the lack of independent objective criteria for the initial selection of the items that constitute the tests.

The mathematical apparatus of factor analysis tends to obscure for some the fundamental logic of factor analytic investigations. The apparatus has been developed to handle simultaneously great quantities of interrelated data representing responses of individuals to test items. The completed analysis, if successful, indicates classes of test items that have elicited, within classes, similar responses from each individual in the sample, but on which similar responses have differed from individual to individual. The several empirically identified classes of items (stimuli) are then given names (e.g., "perceptual speed test," "number test," "test Y," etc.), and individuals receiving high scores on these classes are said to be high in "perceptual speed ability," in "number ability," etc. The prediction can be made that individuals from the appropriate population will tend to behave with intraindividual consistency on items within a class and will differ from each other in the consistent mode of behavior within classes of items, and that relatively little consistency in behavior will be manifested from class to class. One of the presumed goals of this procedure is that tasks other than those used previously will yield to an objective analysis which will permit one to specify the combination of scores on the isolated factors that will be appropriate for successful performance on the task. Explicit rules for such analyses are not yet available. If such rules are ever specified, the utility of this approach will have been demonstrated.

Experimentation using factor analysis has attempted to study simultaneously groups of items toward which individuals behave with intraindividual consistency and with individual differences in the manner of responding to these classes of items. Except for the latter problem, the procedure does not differ in fundamental logic from the procedures that have been used to scale the psychological similarity of stimulus items. The meaning of the term "number test" or any other test can be given by stating the criteria for classifying the items into the test; this includes the entire factor analytic procedure. The meaning of the term "numerical ability" is given when the test is specified, the rules for administering it are given, and the scoring criteria stated. The term "factor" has often been used to refer to these different, though related, concepts. More "operational" definitions of psychological concepts could scarcely be given. It should be clear, however, that no "primary" factors, in the sense of physiologically or phenomenologically fundamental variables, can be said to have been isolated by the procedure utilized by the factor analyst any more than this could be said of other definitional procedures in psychology.

There is little sense to the question: "Which of these definitions of intelligence is correct (or most nearly correct)?" *Formally* correct definitions of all these concepts may be given. Which of the several concepts of intelligence proves to be the most useful, in the sense of entering into laws which lead ultimately to more accurate predictions of human behavior, remains to be seen. There is little use in speculating unduly on this point, considering our current state of ignorance concerning the variables associated with these concepts. Only empirical research can provide an unequivocal answer.

A similar analysis clarifies arguments concerning whether or not intelligence tests *need* to contain "nonintellective" items. We may recognize, first, that the occurrence of the terms "intellective" and "nonintellective" in everyday language does not guarantee that they refer to any features or phenomena that may be either consistently or usefully distinguished. If it is assumed for the moment that the terms are both useful and unambiguous, the proper question to ask is whether or not such items in a test will facilitate the achievement of the purpose for which the test was constructed. Test constructors are (understandably) rarely explicit about *all* the predictions they wish to make with their tests, and it is impossible to determine, a priori, whether or not any particular class of items will prove generally useful. Many of the controversial points concerning "the nature of intelligence" stem from an assumption that all investigators constructing or working with "intelligence tests" have a single common goal.

In this connection, Wechsler (10) asks if "the capacity for social adaptation" is not also a "sign of intelligence." He states that intelligence tests involve more than "mere learning ability or reasoning ability or even general intellectual ability." They also contain other "capacities which cannot be defined as either purely cognitive or intellective." He goes on to state that this is desirable, and that such factors should be included with greater premeditation. One might well ask how one is to arrive at a sensible decision on this proposal until the goals of intelligence testing have been relatively clearly set forth. The issue, it would seem, is not one of a definition of an "absolute" intelligence that will be used generally; rather, it is necessary to state explicitly the criterion (or criteria) to be predicted, and then to discover the tasks that will predict it.

Heredity-environment. One of the most intense controversies in psychology in recent years was the heredity-environment issue. On the one side[1] was a group of individuals insisting that "intelligence" is something not directly influenced by the environment, i.e., not directly influenced by learning. On the other side, it was maintained that intelligence could be affected by learning experiences. This issue was closely related to the argument over the constancy of IQ, the insistence that IQ's obtained from certain tests (viz., the Stanford-Binet) did or did not fluctuate markedly from time to time for a given individual. Reverberations of these controversies are still heard in current discussions of culture-free intelligence tests.

The salient points in this controversy were rarely, if ever, clearly and explicitly delineated. The polemical papers written on the subject indicate that much of the difficulty centered around careless use of terminology on both sides, and they suggest that a methodological analysis should prove clarifying. For example, the terms "environment" and "heredity" were never clearly de-

[1] The writers know of no reputable psychologist who could be said to belong unequivocally in one or the other of these mythical groups. Rather, the points at issue have been schematized in this way in order to represent more simply the pattern of the controversy.

fined, thus sharing the same ambiguity as "intelligence"—the concept they were intended to clarify. In the biological sciences, the term "heredity" is used precisely only in relation to the genotypically traced characteristics of the ancestors of the individual whose heredity is under discussion. Any attempts to define "intelligence" by referring to "heredity" would presuppose application of the procedures of the geneticist to the "intelligence" of the ancestry—and the circularity of this is apparent.

When one turns to research on the relationships between "heredity" and "environment" on the one hand and "intelligence" on the other, and construes these concepts operationally in terms of the research reports, one finds numerous definitions. A typical pattern of research was to provide an experimental group of children with specified experiences, to give pre- and posttraining intelligence tests, and then to compare the IQ gains with those of a control group not having the same intervening training. If greater gains occurred for the experimental group than for the control, it was held that the "environment" had influenced "intelligence." Few, if any, of these studies were devoid of serious experimental errors, the most damaging of which, in the writers' opinion, was the typical failure to assign subjects at random to the experimental and control groups. The foster home studies provide another pattern of research used by the "environmentalists," and were similarly limited by experimental errors.

The "hereditarians" had their own crucial experimental designs. If the IQ's for pairs of siblings reared separately correlated positively and significantly, it was the result of common heredity. If the IQ's for pairs of monozygotic twins correlated significantly higher than the IQ's for pairs of bizygotic twins, it was the result of more similar heredity for the former. Questions arise as to the importance of common uterine experiences, of the physical similarity of identical twins in leading to more similar environmental experiences, of the reliability in identifying identical twins except at birth, and so on. Jones (9) includes critical analyses of many papers in this area.

Much of the argument on the heredity-environment issue was not confined to such empirical questions as the foregoing paragraphs describe. Many workers in the area desired and expected a concept of intelligence which would provide a quantitative index that would not change with time for the individual except under the most unusual conditions, e.g., brain damage, psychosis, paralysis, etc. An intelligence test which suggested that intelligence fluctuated from day to day was therefore unsatisfactory; it was not a "real measure" of intelligence. The first empirical studies reporting systematic changes in IQ for groups were looked upon with considerable suspicion by many investigators. These studies and their supporters were answered with suggestions about uncontrolled variables that might have produced changes in IQ scores without affecting the fundamental intelligence. It now appears that this objection referred to the plausible possibility that IQ scores may be changed without materially affecting performances on tasks for which there was either a presumed or an experimentally established relationship with the IQ scores. The literature shows an interesting neglect of this possibility by those who insisted on the effectiveness of environmental factors in changing the level of intelligence. An obvious example of such a factor is coaching.

A terminological analysis helps to bring the conflicting conclusions into agreement. If intelligence is understood to refer to the performance on a given scale (Meaning I only), then without question, some environmental influences (e.g., coaching, repetition of tests, etc.) can produce changes in intelligence. On the other hand, if intelligence is understood to refer to some complex set of interrelated behaviors (Meaning I *and* Meaning II), and if we have neither a complete list of the behaviors nor explicit statements of the relations holding among them, then we do not know and cannot determine whether or not learning experiences can produce changes in intelligence. As a matter of fact, if intelligence is understood in this sense, we can never know fully what intelligence "means," since subsequent investigations may uncover new relationships between the behavior and other concepts. One of the more important results of a methodological analysis of a scientific concept is the distinction made between the formal meaning of the concept and the empirical knowledge about the concept.

Analysis of the heredity-environment issue cannot be considered complete until mention has been made of the scientifically irrelevant values that have still further clouded the issues involved. The common-sense meaning of "intelligence" has a high value connotation for most of us, a characteristic it shares with many other psychological concepts (e.g., "rigidity," "neurosis," "prejudice," etc.). Intelligence tests have thus been evaluated by some, not only in terms of their predictive power, but also in terms of the "desirability" of the content. The evaluations seem to state: "Intelligence is 'good,' and if the test does not predict 'good' behavior, then it is not an intelligence test." This attitude often results in either a high evaluation of the IQ, per se, without adequate consideration for what can be predicted from it, or in bitter denunciation of test constructors who include questions in the test which handicap certain groups.

To ask whether it is good or bad for an individual to have high intelligence is about as scientifically relevant as to ask whether it is good or bad to have an object weigh a lot. After scientists have defined their terms and have stated the interrelations among them, societies may decide whether or not a given term refers to something desirable. To reverse the procedure places on the scientist "pious" restrictions that are irrelevant to his purposes.

A survey of current literature on culture-free intelligence tests demonstrates this confusion of value and factual matters. For example, Eells *et al.* (7), with the most articulate of frames of reference, criticize the modern educational system and, therefore, the intelligence tests that predict success in it. They point out that middle-class teachers, with their particular middle-class version of what is the "best" and "true" culture, inflict their values upon school curricula, judgments of their pupils, and intelligence test items. Thus, they fail to develop the "full mental capacities" of their pupils, particularly of those pupils from lower classes. Present intelligence tests seek to predict behavior closely related to the school culture. They are, therefore, inadequate "to measure the general problem-solving activities of human beings." What is needed is an intelligence test that reflects or measures the "genetic mental equipment," "the general problem-solving activities," "the real talents," etc. Such an index would permit us to show that class differences in intelligence do not exist and thus help to prevent social class prejudice and untoward discrimination.

Without arguing for or against the educational goals of Eells and his coworkers, we make the following comments. Most psychologists would now agree that the predictive power of intelligence tests has been grossly overestimated, in both scope and accuracy, by many professional and nonprofessional people. But to criticize a test because it predicts one thing and not another seems pointless. Whether or not a test can be constructed to predict important behavior, and yet not discriminate among social classes, is entirely a question of fact. Apparently Eells *et al.* (7) are attempting to construct such a test, and their attempts to make explicit the behavior they consider it important to predict should aid them. That part of their program concerned with a reformulation of educational goals can find no direct support from scientific knowledge since science cannot tell us what the "better life" is.

The validity of intelligence tests. Attempts to use technically the ambiguous term "validity" have generated much confusion in literature on intelligence. Consider the basic question, "Is this intelligence test valid?" One possible clear meaning of this vague question has to do with the usefulness of the test for predictive purposes. The answer to the question, by this interpretation, requires only a summary of the empirical research with the test. There is, of course, not much point in asking the question about a new test since little empirical knowledge will be available. If a new test is demonstrated to predict the scores on an older, well-established test, then the evaluation of the predictive power of the older test may be used for the new one. In this sense, the "validity" of a new test may be established relatively easily. Usually, however, the publication of a new test should be regarded as an invitation for other investigators to help to discover the predictive power of the test. If a given investigator judges that claims are made for the test that are not warranted by the empirical data, then it is his duty to register his objections. But a bland statement that the test is not valid contributes nothing but confusion and polemics to psychological knowledge. It amounts to nothing more than a forecast of future uselessness of the test.

The previous interpretation of the basic question has the virtue of permitting an eventual empirical answer. Another frequent interpretation is not so fortunate, having to do with whether or not the test is a *true* measure of intelligence. It presupposes a meaningful concept of *true intelligence.* It seems that such a question, unanalyzed, has led many workers to attempt to discover the "underlying nature of intelligence." It is rarely clear from their writings what is the "nature" of the "nature" they expect to find. It appears to have something to do either with the physiology or with the mental data of their subjects. The comments that follow are devoted to the issues that seem to be involved.

If one defines "intelligence" (or any other psychological concept) in terms of the individual's responses to items on a standardized test, one may still ask, "What are the physiological correlates of this type of behavior?" That every bit of behavior has physiological correlates is something of which psychologists are, as Bergmann puts it, "as certain of as we are of anything in science" (3, p. 442). Unfortunately, the more complex (molar) the behavior, the more likely it is that our present best attempts to specify which physiological variables underlie the behavior will be pure speculation and probably will be neither good psychology nor good physiology.

The problem is not greatly different in practice if one asks, "What are the mental correlates of this type of behavior?" No psychologist claims direct observation of his subject's mental data. If he is to do more than speculate, he must settle for observation of the subject's behavior (including verbal behavior) and the situations in which it occurs. He must assume that no mental states occur which are not *in some way* reflected in observable behavior.

The only important point that needs

to be made is that both the mental and the physiological correlates remain forever distinct from the behaviorally defined (psychological) concepts. Even if one finds an invariant relationship between a psychological and a physiological variable, they remain two things. One has found a law relating them. The failure to recognize this point has apparently led some writers to think of the physiological or mental variables as the "true" ones, which are only approximately "measured" by behavioral variables. What some psychologists seem to ask is whether or not the test reflects accurately the appropriate mental variables. The hopelessness of any immediate attempt to answer such a question is obvious. The most convincing answer one could give is the same answer one would give to the question, "How adequately does the test predict certain areas of behavior?"

To avoid misunderstanding, it should be made explicit that this formulation does not suggest that the study of the relationship between psychological and physiological variables is either an illegitimate or an unprofitable area for psychologists. Nor does it suggest that the study of subjects' verbal responses, under special instructional sets and conditions, as they relate to other situational or behavioral variables, is either a logical or factual error. The argument is merely that there are no a priori reasons why these variables are more fundamental ("real") than those at the behavioral level. This is a matter to be determined only by empirical trial and error.

SUMMARY

This paper is an attempt to examine some of the controversial issues in the field of intelligence by an application of some basic principles in the philosophy of science. A summary of the most relevant of these principles was given, and the principles were then applied to such problems as the organization of intelligence, the heredity-environment issue, and the validity of intelligence tests. The aim of the analysis in each case was to separate terminological and other logical issues from the factual issues with which they have become confused. It was seen that there is little left that can be considered controversial, except in the sense that any question of fact may be a controversial point until adequate evidence is provided for its resolution. The confusions that arise as a result of trying to formulate single answers to multibarrelled questions can be eliminated.

REFERENCES

1. BECHTOLDT, H. P. Selection. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 1237–1266.
2. BERGMANN, G. The logic of psychological concepts. *Phil. Sci.,* 1951, **18,** 93–110.
3. BERGMANN, G. Theoretical psychology. *Annu. Rev. Psychol.,* 1953, **4,** 435–458.
4. BERGMANN, G., & SPENCE, K. W. The logic of psychophysical measurement. *Psychol. Rev.,* 1944, **51,** 1–24.
5. BRUNSWIK, E. The conceptual framework of psychology. Chicago: Univer. of Chicago Press, 1952. (*Int. Encycl. unified Sci.,* v. 1, no. 10.)
6. CARNAP, R. Testability and meaning. *Phil. Sci.,* 1936, **3,** 418–471; 1937, **4,** 1–40.
7. EELLS, K., DAVIS, A., HAVIGHURST, R., HERRICK, V. E., & TYLER, R. *Intelligence and cultural differences.* Chicago: Univer. of Chicago Press, 1951.
8. FEIGL, H. Operationism and scientific method. *Psychol. Rev.,* 1945, **52,** 250–259.
9. JONES, H. E. Environmental influences on mental development. In L. Carmichael (Ed.), *Manual of child psychology.* New York: Wiley, 1946. Pp. 582–632.
10. WECHSLER, D. *The measurement of adult intelligence.* Baltimore: Williams & Wilkins, 1944.

(Received November 4, 1953)

Psychological Review
Vol. 61, No. 4, 1954

THE SKAGGS–ROBINSON HYPOTHESIS AS AN ARTIFACT OF RESPONSE DEFINITION[1]

MALCOLM L. RITCHIE

University of Illinois

The literature on retroaction shows many variables involved in the determination of the experimental results. One of the most important of these is the similarity between original and interpolated tasks. Systematic investigations of similarity have been conducted since 1920. These experiments show that interference effects have either increased (e.g., 3, 7) or decreased (e.g., 4, 6) with increasing similarity between the tasks. Apparently, no experiment has reported reliable increasing and decreasing effects in the same experimental design.

There have been two major theoretical attempts (5, 6) to integrate the results of the similarity experiments to form a general similarity function. These general functions have assumed that the results of the many experiments are comparable even though they involve (*a*) many different definitions of similarity, and (*b*) more than one experimental design. It will be argued here that a confusion of experimental designs within given experiments has led to a procedural problem which is crucial to the interpretation of similarity results. This problem is a fundamental ambiguity concerning the definition of an acceptable response.

When the problem of response definition is recognized, two important consequences are apparent. First, a central issue in retroaction theory—the similarity paradox—becomes a pseudoproblem. Second, one of the most frequently used measures of similarity—identical elements—is seen to be confounded.

The Similarity Paradox

From early systematic studies of similarity, a generalization was drawn to the effect that interference increases as the similarity of the tasks increases. An extrapolation of this trend would show maximum interference when the tasks are made maximally similar. A paradox arose when it was noted that maximal similarity of successively learned tasks is the condition for ordinary learning; that is, continued learning with the same materials. Thus it appeared that the point of maximal similarity was at once the condition of (*a*) maximum facilitation, and (*b*) maximum interference. This is the similarity paradox to which Robinson (6) called attention and for which he proposed a resolution, which has come to be known as the Skaggs-Robinson hypothesis.

The Skaggs-Robinson resolution. Robinson began his analysis with the experimental data available to him, which showed increasing interference with increasing similarity. But he reasoned that there must be some point along this function at which the trend reverses, interference starts to decrease, and the function moves to high facilitation at the end point of maximum similarity. He also reasoned that there must be at the other end of the similarity scale a point at which no interference or facilitation is found as a function of complete dissimilarity of materials. The curve of this hypo-

[1] This research was supported in part by the United States Air Force under Contract AF 33(038)–25726, monitored by the Air Force Personnel and Training Research Center. Permission is granted for reproduction, translation, publication, use and disposal in whole and in part by or for the United States Government.

thetical function goes from high facilitation (point A on Fig. 1) at maximum similarity to maximum interference at some intermediate point (point B), then to neutral at the point of complete dissimilarity (point C). Robinson had available evidence showing the slope from B to C. With a different experimental design he obtained results (6) giving the slope from A to B. However, many subsequent attempts have failed to produce the entire function within one experimental design and with one definition of similarity.

Osgood's transfer-retroaction theory. Osgood (5) pointed out the lack of specificity of Robinson's formulation. In a stimulus-response analysis of the experimental materials, he showed clearly that three basic types of experiments have been used: (a) stimuli constant and responses varied, (b) responses constant and stimuli varied, and (c) both stimuli and responses varied. The evidence, which Robinson tried to reconcile with ordinary learning (increasing interference with increasing similarity), Osgood holds to be the special case of simultaneous variation of both stimuli and responses. However, within this special case, Osgood's theory requires that the trend reverse itself in order to account for ordinary learning. The function thus obtained is very similar to that of Robinson.

This analysis of the experimental materials appears to leave the fundamental paradox unresolved. One fact that has been overlooked is that there exists within these experiments a confusion of experimental designs. This confounding renders ambiguous the criteria by which the experimenter defines a correct response.

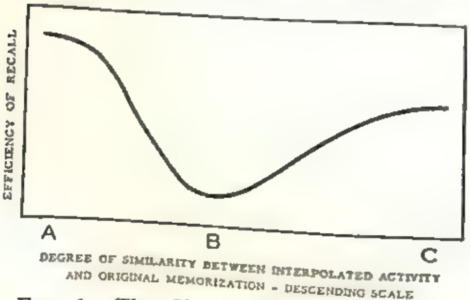

FIG. 1. The Skaggs-Robinson hypothesis. Hypothetical relation between similarity of original and interpolated material and amount of reproductive interference (6).

THE PROBLEM OF RESPONSE DEFINITION

In order for the results of different experiments to be combined in a general function, it must be assumed that the experimental procedures are comparable. When ordinary learning and interference data are compared, this assumption is violated. If we consider the criteria by which the experimenter defines a correct response, it can be shown clearly that two different procedures are involved. The procedure of the interference studies may be expressed as an ABA design, that of ordinary learning as an AAA design.

Response definition in the ABA design. The basic experimental design which has been used in the study of similarity effects is expressed in the ABA paradigm. Three learning series are involved: original learning (OL), interpolation of a different learning task (IL), and relearning (or recalling) the material of the original series (RL). The similarity between the original and the interpolated tasks is varied systematically. In the relearning series the subject is required to make the responses appropriate to the A series and not to make the responses of the B series. One of the stimulus-response paradigms for this design is as follows:

OL	IL	RL
S_1—R_1	S_1—R_2	S_1—R_1

The stimuli are constant for the three series and the responses are varied. It will be noted in this design that the subject must learn two different responses to functionally identical[2] stimuli. The

[2] The point has been made (5) that no two stimuli are ever absolutely identical. We may use the term "functional identity" to express our constant stimulus presentation.

performance on the relearning series involves a competition between the two responses. The subject must make R_1 and must not make R_2 in order for the trial to be recorded as correct. Regardless of how similar the two responses are, the subject is required to discriminate between them in order to satisfy the criteria of a correct response.

Now let us suppose that the experimenter increases the similarity of the required R_1 and R_2 toward the point at which they are functionally identical. As this point is approached, the subject has increasing difficulty in discriminating between them (the experimenter must retain a method of ready distinction), and the interference effects would thus be expected to increase. At functional identity the discrimination cannot be made. If the procedure of the interference design (ABA) is maintained, the interference effects between the two tasks are maximum—the subject cannot learn the two tasks at all.

Response definition in the AAA design. Ordinary learning proceeds on the basis of the successive presentation of stimuli and responses of functional identity. If we set up a paradigm for ordinary learning in the same manner as we have done for the ABA design above, our responses are distinguished only by the series in which they appear.

OL	IL	RL
$S_1—R_1$	$S_1—R_2$	$S_1—R_1$ or R_2

In this case R_1 and R_2 are functionally identical and are distinguished only by the series in which they appear. The subject is not required to discriminate between them in his performance. Either R_1 or R_2 is recorded by the experimenter as a correct response. This means that the subject does not have to discriminate between the response he made in OL and the response of IL. In this situation, facilitation is maximum at the same point as described above—that of functional identity of both stimuli and responses. The difference between the two is the procedure by which the experimenter determines what is to be recorded as a correct response.

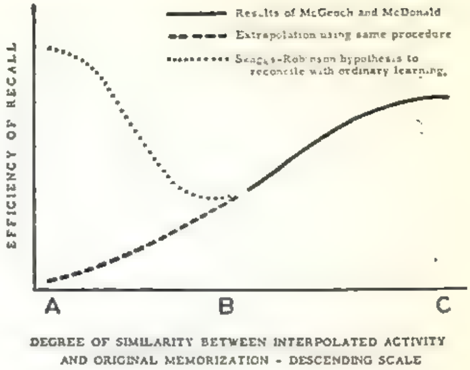

FIG. 2. Diagram showing results of McGeoch and McDonald (3), extrapolation involving continued use of ABA procedure, and the Skaggs-Robinson proposal to reconcile results with ordinary learning.

Robinson's analysis. The data that were available to Robinson were based upon the ABA procedure and showed increasing interference as similarity increased. The results of McGeoch and McDonald (3) have been used in Fig. 2 to represent this trend. Extrapolation of these results would give a prediction of maximum interference at maximum similarity. This extrapolation is based upon continued use of the ABA procedure and is shown by the dash line in Fig. 2. Robinson reasoned that there must be a reversal of this trend in order to account for ordinary learning (dotted line). Robinson's hypothesis requires a shift to the AAA procedure. The writer knows of no results in which this reversal occurred reliably when the same procedure was maintained.

METHODOLOGY AND THE DEFINITION OF SIMILARITY

From Fig. 2 it appears that the similarity function for the AAA procedure is considerably different from that for the ABA procedure. It follows that any experimental design should involve only one of these for the results to be meaningful. The inclusion of both AAA and ABA in a given similarity experiment would confound the effects due to similarity with the effects due to the use of the two procedures. This is just the confounding that occurs when similarity is measured by the number of identical elements.

In the experiment of Robinson (6), for example, similarity was measured by the number of elements that were identical in the three series. The responses in his tasks have been schematized as follows:

			Elements
OL	IL	RL	in Common
a b c d	e f g h	a b c d	none
a b c d	*a* f g h	a b c d	one
a b c d	*a b* g h	a b c d	two
a b c d	*a b c* h	a b c d	three
a b c d	*a b c d*	a b c d	four

From the foregoing analysis it will be seen that wherever a response element appears in all three learning series, the AAA procedure is followed with respect to that element; that is, the subject is not required to discriminate between R_1 and R_2. Wherever a different element is found in the interpolated task, the ABA procedure is followed; that is, only R_1 is acceptable as a correct response in relearning. Thus, in Robinson's experiment, increasing similarity was accompanied by an increase in the use of the facilitative AAA procedure. The less the interference procedure is used, the less would interference effects be expected in the results. This is precisely the result obtained by Robinson (6), Harden (1), Kennelly (2), and others using identical elements as a measure of similarity.

If varying the number of identical elements involves confounding similarity with the differential use of two procedures, then identical elements experiments do not give a legitimate similarity function. Experimental techniques may possibly be devised to separate the similarity effects from the procedural effects. Until this is done, the results of identical elements experiments cannot be compared with those in which only one procedure is used.

Summary

It has been maintained in this discussion that the similarity paradox in human learning was created by the analysis made by Robinson. Maximum similarity between OL and IL may be the condition for either maximum facilitation or maximum interference, depending upon the criteria established by the experimenter for defining a correct response in RL.

The evidence used by Robinson to support his general function is based upon an identical elements definition of similarity. Varying the number of identical elements was shown to confound similarity with the use of two response definitions.

REFERENCES

1. Harden, L. M. A quantitative study of the similarity factor in retroactive inhibition. *J. gen. Psychol.*, 1929, 2, 421–432.
2. Kennelly, T. W. The role of similarity in retroactive inhibition. *Arch. Psychol., N. Y.*, 1941, 37, No. 260.
3. McGeoch, J. A., & McDonald, W. T. Meaningful relation and retroactive inhibition. *Amer. J. Psychol.*, 1931, 43, 579–588.
4. Osgood, C. E. Meaningful similarity and interference in learning. *J. exp. Psychol.*, 1946, 36, 277–301.
5. Osgood, C. E. The similarity paradox in human learning: A resolution. *Psychol. Rev.*, 1949, 56, 132–143.
6. Robinson, E. S. The 'similarity' factor in retroaction. *Amer. J. Psychol.*, 1927, 39, 297–312.
7. Sisson, E. D. Retroactive inhibition: The influence of the degree of associative value of original and interpolated list. *J. exp. Psychol.*, 1938, 22, 573–580.

(Received October 19, 1953)

Psychological Review
Vol. 61, No. 4, 1954

FIELD THEORY: II. SOME MATHEMATICAL APPLICATIONS TO COMPARATIVE PSYCHOLOGY

WILLARD E. CALDWELL[1]

The George Washington University

Until vast amounts of research are carried out on animals other than the white rat and using drives other than hunger and thirst, and until all the quantitative interrelations are investigated, a field theory for comparative psychology will be relatively immature.

The concept of field has had many different applications in both physics and psychology. The attempt will be made here to point out some aspects of this concept as it is being used in this paper. Reference may be made to two previous articles developing this general frame of reference (1, 2). The organism as conceived here is a configuration of energy existing within a larger configuration of energy termed the environment. There is a constant interaction between the two fields. The organism is conceived of as following a process of differentiation with respect to the environmental field. This process reflects an attempt of the organism to achieve homeostasis with the environmental field. Examining a brightly lighted maze and a dark goal box may be viewed as a field of energy with a difference, and this difference may be measured by the intensity of light in the maze and in the goal box. The frame of reference used here is that the organism differentiates in the direction of less light, and in so doing it differentiates between the correct pathway and the incorrect pathway. In essence, then, a difference in the environmental field of energy produces a difference in the organism's field, as manifested in terms of time and errors. The field concept is utilized here in a broader sense, inasmuch as the light is only one part of the total field of energy operating upon the organism; and we shall attempt to present a theoretical framework for isolating some of the major parts of the environmental field with respect to learning and perceptual problems.

In an earlier paper (3) the writer outlined a theory for, and some psychophysical techniques of, investigating maze learning when many different field-type drives are utilized. The concept of field drive is utilized somewhat synonymously with the term exteroceptive stimulus. It covers stimuli which originate outside the organism such as light, temperature, etc. This concept is used to cover types of motivation other than the internal biological drives such as hunger and thirst.

The first mathematical postulate in the earlier paper pertained to an experiment in which temperature and its reduction served as the motivating factor and as reinforcement. It dealt with the ratio between the increment of the difference between temperature in maze and goal box and the difference between temperature in maze and goal box necessary to produce a statistically significant difference in time and errors.

The object of this paper is to illustrate further how this psychophysical approach might be applied to maze learning, motivation, perceptual discrimination, and the Skinner-type design. The attempt will be made to illustrate it as an operational frame of reference applicable to exteroceptive types of stimulation and to many different types of animals.

[1] The author wishes to express his appreciation to Mary Lou Krehbiel for editorial assistance in the preparation of this paper.

TABLE 1

POSSIBLE COMBINATIONS OF EXTEROCEPTIVE STIMULI WHICH COULD THEORETICALLY BE TESTED PSYCHOPHYSICALLY IN MAZE, PERCEPTUAL DISCRIMINATION, AND SKINNER–TYPE DESIGNS

Apparatus (with modifications) and Locations of Measure of the Field	Fields of Energy	Formula for Varying both Measures of the Energy Field Concomitantly	Formula for Varying Stimulus in the Maze, in the Entrance Compartment, and before Pressing Bar	Formula for Varying Stimulus in the Goal Box or the Period after Pressing the Bar
Maze 1. Maze 2. Goal box	Temperature Light Revolutions Gaseous formaldehyde Angle of inclination Sucrose Amperage Sound Humidity	$\Delta(M-g)=K(M-g)$ M =Maze g =Goal box	$\Delta(M-g_c)=K(M-g_c)$	$\Delta(M_c-g)=K(M_c-g)$
Perceptual Discrimination 1. Entrance compartment 2. Goal box		$\Delta(E-g)=K(E-g)$ E =Entrance compartment g =Goal box	$\Delta(E-g_c)=K(E-g_c)$	$\Delta(E_c-g)=K(E_c-g)$
Skinner Design 1. Before pressing bar 2. After pressing bar		$\Delta(B-A)=K(B-A)$ B =period before pressing lever A =period after pressing lever	$\Delta(B-A_c)=K(B-A_c)$	$\Delta(B_c-A)=K(B_c-A)$

SOME APPLICATIONS OF THE BASIC POSTULATE TO MAZE LEARNING

Table 1 presents an outline illustrating some of the permutations to which the writer's mathematical formulation would be applicable. The first major division of its applicability is maze learning. This postulate stated in abbreviated form is: *The increment of the difference between the stimulus in the maze and the stimulus in the goal box necessary to produce a just noticeable difference (j.n.d.) in time and errors is a constant fraction of the difference between the stimulus in the maze and the stimulus in the goal box.* The formula for this may be found in the third column of Table 1. The goal box and the maze both can vary. These formulae can be found in columns 4 and 5. Some preliminary work has been carried out to set up apparatus and procedures for testing some of the different types of stimulation utilized as motivation in maze learning.

Caldwell and Mosman (5) utilized temperature. Caldwell, Thaler, and Katz (10) performed a variation of the temperature-type experiment. Caldwell and Womack (12) utilized light avoidance. Caldwell and Sandler (9) utilized gaseous formaldehyde. Albino mice were utilized in all five of these experiments.

Caldwell and Floyd (4) performed an experiment on albino mice placed in a maze which could turn a certain number of revolutions per minute and which would stop turning when the animals reached the goal box. This type of design lends itself to the possibility of varying the revolutions in the maze and in the goal box. Caldwell and Richmond (7) performed an experiment on hamsters wherein they utilized geotropism as the motivating factor. The maze had an angle of inclination of 21 degrees. The animals had to ascend the maze to reach the goal box, which had an angle of inclination of zero degrees. This experiment was also repeated by Caldwell and Ostrich (6) utilizing albino mice. This type of design also lends itself to the possibility of testing various combinations of differences represented by variations in the angle of inclination.

It is possible to use fish in field-drive experiments. An experiment was performed on goldfish which swam in a maze of high temperature to a goal box of low temperature (11). In this connection it is interesting to hypothesize an electrical field in a maze and either its absence or reduction in a goal box. The maze itself might be positively charged and the goal box negatively charged. Various degrees of conduction might be applied in each.

The salmon has some photosensitive receptors deep in its skin (15). These are first covered by a layer of pigment, which subsequently disappears. As a result of this the fish reacts negatively to light. The hypothesis formulated here is that these receptors in the salmon could operate as motivating factors in the maze, and their reduction would serve as reinforcement. In testing this, various kinds of controls would have to be employed to separate the skin receptors from those of the eye.

Guttman (14) conducted some interesting experiments on rats in which he used bar-pressing responses reinforced with sucrose solutions of various combinations. This suggests the question of how much increment in sucrose is necessary to get a difference in rate of responding in the maze situation. The problem might be stated as the increment of the difference between the sucrose concentration fed before starting the maze and that fed in the goal box necessary to produce a j.n.d. in time and errors representing a constant fraction of the difference between the concentration fed before starting the maze and concentration fed in the goal box.

The stimulation of sound and humidity might be applied to this type of design and the various combinations of differences tested.

SOME APPLICATIONS TO PERCEPTION PROBLEMS

In Table 1 the same approach might be applicable to problems of perception. Flynn and Jerome's study (13) with rats might be applicable here. Light avoidance has been utilized with pigeons on perception problems (8). The light avoidance was employed for motivation in training pigeons to discriminate geometrical figures. The apparatus consisted of a box which was brightly lighted and painted with aluminum paint. The goal box was relatively dark, and the goal-box doors were a circle or a triangle, depending upon the problem. If a pigeon entered the circle, for instance, the light would be turned off and the bird would remain in the dark goal box for five minutes. Error and time curves were established for these pigeons.

Our intention is to emphasize the possible applicability not only of light avoidance but of other types of field drives to problems of perceptual discrimination, and also to urge the psychophysical treatment of data derived from such types of experiments. The part of the apparatus where the animal is placed to make the discrimination is referred to as the entrance compartment, and the darkened area is designated as the goal box. The stimulation can vary in three ways similar to those suggested for maze learning.

For purposes of clarification, this problem might be stated as follows: *The increment of the difference between the light in the entrance compartment and the light in the goal box necessary to produce a j.n.d. in perceptual differentiation (in time and errors and correct choices) is a constant fraction of the difference between the light of the entrance compartment and that of the goal box.*

SOME APPLICATIONS TO THE SKINNER-TYPE DESIGN

Another problem is that of using field drives such as temperature, gaseous formaldehyde, light avoidance, etc. in the type of experimental design outlined by Skinner (15). Skinner's method may possibly be more sensitive to psychophysical measures than those of the standard maze-learning phenomena. Also, it is important theoretically to know how the results obtained from utilizing the Skinner-type design and field drives compare with results derived from the use of maze-learning designs and field drives.

In order to test the following drives, the Skinner-type design must be modified, but the essential elements in the design should be retained—mainly, the instrumental one where a stimulus is introduced and the organism presses a bar to stop the stimulus. Records should be kept on the relation between the variation in the intensity of the stimulus and the variation in bar pressing. Guttman investigated bar-pressing responses in the rat where sucrose was utilized as reinforcement. He says:

> Evidence is presented that rate of responding in the Skinner box with rats is a semilogarithmic function of the concentration of sucrose used as reinforcement. Extrapolation of the fitted rate-concentration function yields an estimated reinforcement threshold in the region of the sucrose-preference threshold and the human sucrose limen. Extension of this experimental technique to other reinforcing agents may yield a systematic pattern among reinforcement thresholds (14, pp. 360–361).

The application of this approach to the Skinner design might be further clarified by the following: *The increment of the difference between the intensity of light in the Skinner bar-pressing apparatus before the bar is pressed and the intensity of light there after the bar is pressed necessary to produce a j.n.d. in the time and frequency of bar pressings is a constant fraction of the difference between the intensity in the Skinner bar-pressing apparatus before the bar is pressed and the intensity there after it is pressed.*

Some Applications to Motivation

A fourth type of apparatus to which this general theoretical approach might be applied is that which attempts to measure motivation. There are many types of apparatus which are utilized for measuring activity levels. The revolving drum is one that might be applied here to these various exteroceptive stimuli. The difference in the stimulus field could be measured with light, as an example, by measuring the light intensity in the animal's cage and then measuring it in the revolving drum. The j.n.d.'s would be in terms of activity level measured in terms of the number of revolutions of the revolving drum. This also raises the question of measuring the animal's frame of reference before placing it in the other types of designs mentioned in this paper. This may be one of the advantages of utilizing exteroceptive stimulation rather than hunger or thirst in animal experiments.

Discussion

The foregoing programmatic outline of research is presented in broad outline form. The j.n.d. is actually a statistically significant difference in time, errors, and correct choices. Different parts of these curves obtained should be compared statistically. The quantitative results expected might appear only in experiments with certain animals. Perhaps only certain drives will be of use from a psychophysical point of view, possibly in connection with only a few types of animals, but ascertaining such facts requires that many animals be utilized in testing each variation of the hypothesis.

It may be that many of the experimental designs given here are too variable. The Skinner-type design was referred to for use in testing some hypotheses, but perhaps additional apparatus, more sensitive and of a new type, should be devised. Certainly the apparatus suggested here should be modified for the different species and for measuring such stimuli as light in comparison with the more conventional types of motivating stimuli such as hunger and thirst. Control groups are necessary where there is no difference between the entrance box and the goal box (or its equivalent) with respect to the particular drive being tested.

Details for investigating these abbreviated hypotheses must be worked out for each experiment. Reference should be made to the writer's previous paper (3) on the application of psychophysics to learning and reinforcement.

SUMMARY AND IMPLICATIONS

Historically, the communication of ideas in a form in which the experimentalist can investigate them in the laboratory has been one of the principal functions of psychological theories. There are dangers and limitations in miniature quantitative theories, but there is also value in operationally defining problems so they may yield data that can invalidate or substantiate the basic assumptions underlying a theory.

This paper has urged that, from the psychophysical point of view, a tremendous amount of research is needed in the field of comparative psychology before we can begin to construct theories that possess any degree of maturity, either qualitatively or quantitatively. It also has attempted to present research problems that might be quantitatively tested in connection with the comparative aspects of motivation and reinforcement.

The implications are that further integration of the field drives with perceptual-type experiments, maze learning, motivation, and utilization of more sensitive techniques similar to Skinner's should be attempted in such a way that they may: (*a*) be checked psychophysically, (*b*) be checked with many different species, (*c*) have their functions checked against results obtained from the more conventional maze-learning experiments, and (*d*) yield results which might aid in giving us a more unified field theory for comparative psychology.

REFERENCES

1. CALDWELL, W. E. Adaptive conditioning: a unified theory proposed for conditioning. *J. genet. Psychol.*, 1951, 78, 3–37.
2. CALDWELL, W. E. The theory of adaptive differentiation. *J. Psychol.*, 1951, 31, 105–119.
3. CALDWELL, W. E. The mathematical formulation of a unified field theory. *Psychol. Rev.*, 1953, 60, 64–72.
4. CALDWELL, W. E., & FLOYD, J. P. The performance of albino mice in the maze situation with stimulation of the vestibular sense as motivation and its relative absence as reinforcement. *J. genet. Psychol.*, in press.
5. CALDWELL, W. E., & MOSMAN, K. F. The role of temperature change as reinforcement. *J. Psychol.*, 1951, 32, 231–239.
6. CALDWELL, W. E., & OSTRICH, R. The performance of albino mice in the maze situation utilizing gravitation and the vestibular sense as motivation. *J. genet. Psychol.*, in press.
7. CALDWELL, W. E., & RICHMOND, R. G. The performance of hamsters in the maze situation utilizing gravitation and the vestibular sense as motivation. *J. genet. Psychol.*, in press.
8. CALDWELL, W. E., & RICHMOND, R. G. The utilization of light avoidance as motivation in the investigation of perceptual differentiation in the pigeon. *J. genet. Psychol.*, in press.
9. CALDWELL, W. E., & SANDLER, H. M. The role of gaseous formaldehyde as reinforcement in maze learning in albino mice. *J. Psychol.*, 1952, 33, 47–56.
10. CALDWELL, W. E., THALER, W. D., & KATZ, J. J. The utilization of temperature change as motivation and reinforcement in the maze performance of albino mice. *J. genet. Psychol.*, in press.
11. CALDWELL, W. E., & TIEDEMANN, J. G. The performance of goldfish in the maze situation with the utilization of temperature as motivation and its reduction as reinforcement. *J. genet. Psychol.*, in press.
12. CALDWELL, W. E., & WOMACK, H. The performance of albino mice in the maze situation with the utilization of light as motivation and its relative absence as reinforcement. *J. Psychol.*, 1953, 35, 353–360.
13. FLYNN, J. P., & JEROME, E. A. Learning in an automatic multiple-choice box with light as incentive. *J. comp. physiol. Psychol.*, 1952, 45, 336–340.
14. GUTTMAN, N. Theories of reinforcement and the reinforcement threshold. *Amer. Psychologist*, 1953, 8, 360–361. (Abstract)
15. ROULE, L. *Fishes: their journeys and migrations.* New York: Norton, 1933.
16. SKINNER, B. F. *The behavior of organisms.* New York: Appleton-Century, 1938.

(Received October 10, 1953)

Psychological Review
Vol. 61, No. 4, 1954

CRITICAL COMMENT ON "LEARNING AND THE PRINCIPLE OF INVERSE PROBABILITY"

ROBERT P. ABELSON

Yale University

An impression that the theorems of inverse probability are of widespread applicability to learning theory is created by David Bakan in his recent paper "Learning and the Principle of Inverse Probability" (1). His treatment is very simple and very ingenious and, if one were not to give the matter too much thought, his conclusions would seem quite sound and powerful. Many statements about learning rates, extinction rates, trial-and-error learning, insightful learning, etc. are made in Bakan's paper. The writer has no quarrel with these statements as such; however, it is felt that Bakan's statements do not follow from his premises. The subsequent discussion will attempt to show that inverse probability is not especially cogent to learning theory and that its use in that context is a misrepresentation either of inverse probability or of learning theory.

Bakan defines three entities: g, h, and x. For simplicity's sake, our discussion will include only g and x. However, it should be understood that h, which represents the ability level and prior experience of the organism, is assumed to be known in all of the subsequent definitions. The symbol g is defined as a certain state of the organism, presumably a state in which the organism is capable of responding in a particular way—the organism, when in the state g, might be said to be "knowledgeable." The symbol x is defined as a particular proposition of knowledge, a hypothesis about the environment (e.g., "If I press the bar, I will get a pellet of food").

Bakan's basic equation requires the following definitions:

$P(g)$ is the probability that the organism is in the condition g.

$P(g/x)$ is the probability that the organism is in the condition g after x is verified or reinforced.

$P(x/g)$ is the probability that x will occur if the organism is in the condition g.

$P(x/\bar{g})$ is the probability that x will occur if the organism is *not* in the condition g.

R is the ratio $\dfrac{P(x/g)}{P(x/\bar{g})}$.

Then Bakan writes (legitimately so):

$$P(g/x) = \frac{R\,P(g)}{R\,P(g) + [1 - P(g)]}.$$

From this equation, Bakan derives all his results.

This equation is said to involve inverse probability because it attempts to infer causes from the observation of effects. The equation contains expressions for the probabilities of occurrences which are never observable [$P(g/x)$ and $P(g)$]. These probabilities ordinarily cannot be verified by counting the relative frequencies of favorable occurrences. Indeed, the philosophical dispute from which Bakan takes great pains to dissociate himself is concerned with the question of whether there exists any sense at all in which $P(g/x)$ can be considered a probability.[1] The negative position, tersely stated, is that "either the organism is in the condition g, or it isn't. A probability statement

[1] Carnap (3) is engaged in an extensive logical analysis of probabilistic statements. His work may provide a resolution of the long-standing philosophical dilemma.

is inappropriate." According to this position, many classes of phenomena would be excluded from the realm of discourse of probability. For instance, one would never speak of the probability of the truth of Weber's Law, or of the mass-energy relation $E=mc^2$, or of Freud's theory of unconscious motivation. These laws or theories are (provided the context of their application is defined) either true or they are false. Nor is there any connection between "approximately true" and "moderately probable" insofar as theories or laws are concerned.

From the point of view of the statistician, the only appropriate use of probability with these classes of phenomena is to fall back on the statement: "If I claim this theory to be true, the probability that my claim will be proven correct is such-and-so." Then if many claims are made, one can calculate the expected number of correct claims. This procedure can lead to a maximization policy with respect to claims about unobservable causes—a kind of static "game theory" approach to the collection of knowledge on the nature of the universe. "Maximum likelihood estimation," together with the method of "confidence intervals" in statistical theory (4, pp. 507–513), is such an approach.

Regardless of the philosophical merits of the argument against the use of inverse probability, its cogency in the case of Bakan's derivations is apparent. The argument is particularly devastating when it is realized that the only reasonable interpretation of Bakan's learning curves is that they are functions describing the experimenter's "game." If we take seriously Bakan's formulation, which allows only two conditions for the organism, g and "not-g," then we will find that the learning curve for an individual organism is *not* a gradually rising curve. If we know which condition the organism is in at every trial, we will find that the learning curve approximates a step function. When the organism is in the state $\bar{g}$, it will respond at a low probability level of success, $P(x/\bar{g})$, and will keep responding at this low level until suddenly it attains the state g. At this point, the organism will abruptly start responding with a high probability level of success $P(x/g)$ and will forever after maintain this high level. *All* learning would then be "insightful." Now, conceivably, Bakan's "learning curve" can be viewed as an average of an infinite number of such step functions, each with a different time of cross-over from $\bar{g}$ to g. However, such a "learning curve" would have no meaning when applied to a single organism. To apply Bakan's gradually rising curve to a single organism is simply to admit our ignorance of the *actual* state of the organism at any given time.

Bakan exposes himself all the more to this criticism by using $R = \frac{P(x/g)}{P(x/\bar{g})}$ as a parameter of the learning curve, implying that in a given case one might measure both $P(x/g)$ and $P(x/\bar{g})$ in order to be able to specify the exact form of the learning curve. However, such a measurement can only be made if there is some criterion by which we can determine whether an organism is in g or in $\bar{g}$ so that the relative probabilities of the occurrence of x in these two circumstances can be determined experimentally. But if such a criterion existed, then the sensible thing to do would be to apply it to the organism while it was learning to find out when it was in g and when it was in $\bar{g}$, and thus solve the problem at once. R, then, is a parameter with the following properties:

Either 1. It can never be measured in practice

Or 2. Its measurement destroys the theoretical grounds on which it is based.

Bakan's learning theory is not a theory of the learning process in a given organism; it is a theory of the process of analyzing the learning process of an organism. As such, it is typified by the situation in which a scientist analyzes the advancing state of his own knowledge. It does apply to "the method of science as a way of learning," as Bakan claims, but it applies only in this situation and not to classical learning theory. A rat is certainly not capable of analyzing his own learning process in this complicated way. To extend the results to classical learning theory would constitute a gross misunderstanding.

The current mathematical models for the learning process (Mosteller and Bush [2]; Estes [5]) *put the variable ignorance into the organism itself,* instead of into the experimenter. They assume not two states of the organism, g and $\bar{g}$, but a continuum of possible states p, where p is the probability that the organism will make the correct response. The hypothetical learning curve is given by p as a function of the number of trials. In this type of model, *the organism itself* gradually becomes more and more certain of the correct response. In Bakan's model, the *experimenter* gradually becomes more and more certain that the organism is cognizant of the correct response. The former seems to be much the more appropriate model for the typical learning situation. That is not to say, of course, that Bakan's results cannot prove to be of value in the limited context of scientific method as the experimenter's way of learning.

REFERENCES

1. BAKAN, D. Learning and the principle of inverse probability. *Psychol. Rev.*, 1953, 60, 360–370.
2. BUSH, R. R., & MOSTELLER, F. A stochastic model with applications to learning. *Ann. Math. Statist.*, 1953, 24, 559–585.
3. CARNAP, R. *Logical foundations of probability.* Chicago: Univer. of Chicago Press, 1950.
4. CRAMER, H. *Mathematical methods of statistics.* Princeton: Princeton Univer. Press, 1946.
5. ESTES, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, 57, 94–107.

(Received January 8, 1954)

Psychological Review
Vol. 61, No. 4, 1954

A METABOLIC INTERPRETATION OF INDIVIDUAL DIFFERENCES IN FIGURAL AFTEREFFECTS

MICHAEL WERTHEIMER AND NANCY WERTHEIMER

Wesleyan University

A decade ago, Köhler and Wallach (6) studied the effects of prolonged figural stimulation on certain subsequent perceptual test patterns, a phenomenon they called figural aftereffects. To account for these effects they postulated a change in the polarizability of that part of the brain upon which the previous contour had been projected, a process they termed satiation.

It has been repeatedly observed (2; 3, pp. 202, 207; 4, p. 202; 5, pp. 300, 316; 8; etc.) that there are individual differences in the size of such figural aftereffects. If one tentatively accepts Köhler's physiological model (3, 4, 5, 6), then such individual differences must reflect differences in the ease with which a modification in cortical conductivity can be brought about. Such a view implies that figural aftereffects could be used to measure generalized cortical modifiability in an individual. Assuming that this modifiability is characteristic of the entire brain, and not specific to a given area, this leads to the prediction that the size of figural aftereffects in different modalities should be correlated, i.e., a small kinesthetic figural aftereffect indicates low cortical modifiability, which in turn would predict a small visual figural aftereffect. Specifically, (*a*) visual and kinesthetic figural aftereffects measured on a large number of people should show a positive correlation (as also suggested in 4, pp. 196–197) and (*b*) intraindividual changes in visual and kinesthetic figural aftereffects should pursue a parallel course, assuming that an individual's cortical modifiability will change through time.

The satiation theory, as well as the newer statistical theory (7), involves physicochemical alterations in cortical tissue. These could be interpreted as implying metabolic changes. Thus one could argue that a relatively large figural aftereffect reflects high physicochemical modifiability and hence conceivably a high "metabolic efficiency."[1]

Although this latter term is not sufficiently defined, it is adequate to yield some further predictions: (*c*) Size of figural aftereffect should correlate with physiological indicants of metabolic efficiency such as basal metabolic rate, thyroid activity, and such indices of circulatory efficiency as capillary structure. (*d*) It should similarly correlate with behavioral indicants of neural efficiency such as reaction time and ease of simple sensory-motor learning. (*e*) An experimentally induced alteration in metabolism should be reflected in a concomitant change in the size of figural aftereffects. (*f*) Schizophrenics, as a concrete example of a group of subjects with generally low metabolic efficiency (1, 8), should exhibit smaller figural aftereffects than normal subjects.

All the above predictions have been subjected to at least a preliminary empirical test, and all have been essentially confirmed, with the exception of the one concerning simple sensory-mo-

[1] Originally we used the term "metabolic rate," but experimental evidence has shown that this concept seems to be inadequate. In our data, figural aftereffects are maximal in the normal range of metabolic functioning and fall off on either side. This result has tentatively led us to the term "metabolic efficiency."

tor learning, where the evidence was ambiguous.

Although these predictions are the only ones tested thus far, the present interpretation could yield many more, especially in classes *c* and *d* above, e.g., predictions concerning hormonal balance, stress effects, problem solving, and perceptual rigidity. Further, the vagueness of the present formulation has the virtue of making it compatible with any theory of figural aftereffects in which metabolic changes in neural tissue can reasonably be assumed, be the theory in terms of homogeneous conductors (6) or neural elements (7).

REFERENCES

1. HOSKINS, R. G. *The biology of schizophrenia.* New York: Norton, 1946.
2. KLEIN, G. S., & KRECH, D. Cortical conductivity in the brain-injured. *J. Pers.*, 1952, 21, 118–148.
3. KÖHLER, W. Relational determination in perception. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior: the Hixon symposium.* New York: Wiley, 1951. Pp. 200–230.
4. KÖHLER, W., & DINNERSTEIN, D. Figural aftereffects in kinesthesis. In *Miscellanea Psychologica Albert Michotte.* Louvain, 1947. Pp. 196–220.
5. KÖHLER, W., HELD, R., & O'CONNELL, D. N. An investigation of cortical currents. *Proc. Amer. philos. Soc.*, 1952, 96, 290–330.
6. KÖHLER, W., & WALLACH, H. Figural aftereffects: an investigation of visual processes. *Proc. Amer. philos. Soc.*, 1944, 88, 269–357.
7. OSGOOD, C. E., & HEYER, A. W. A new interpretation of figural aftereffects. *Psychol. Rev.*, 1952, 59, 98–118.
8. WERTHEIMER, M. The differential satiability of schizophrenic and normal subjects: a test of a deduction from the theory of figural aftereffects. *J. gen. Psychol.*, in press.

(Received January 4, 1954)

Vol. 61, No. 5 September, 1954

THE PSYCHOLOGICAL REVIEW

THE STRUCTURING OF EVENTS: OUTLINE OF A GENERAL THEORY WITH APPLICATIONS TO PSYCHOLOGY[1]

FLOYD H. ALLPORT

Maxwell Graduate School, Syracuse University

I. The Problem of Structure

It is generally assumed that the way to understand nature lies in the understanding of its laws. Some predictability and order in the objects studied are a prerequisite to knowledge about them. It is frequently assumed, also, that the laws of nature are essentially *quantitative*—that they express amounts of some measurable attribute and relationships by which one such variable is a function of another. Hull and his associates stated the matter as follows: "Since it appears probable that everything which exists at all in nature exists in some amount, it would seem that the ultimate form of all scientific postulates should be quantitative" (2, p. 8).

The purpose of the present article is to explore the possibility that, notwithstanding the ubiquity, precision, and unquestioned importance of quantitative and covariational formulas, there may be in nature *another type* of law that is quite as universal, objectively demonstrable, and, in its way, precise. The writer believes that this is true and that the knowledge of such a possible nonquantitative, but still fundamental, law (which, however, is neither "qualitative" nor "configurational") is as indispensable as quantitative formulations for a full understanding of any phenomenon, and that it is particularly needed at the present time in the field of psychology. It is to knowledge of this sort that we must turn for further illumination upon that still unsolved but vital problem, the organization of behavior. However useful they may be for descriptive purposes, the molar laws of covarying behavioral quantities have about reached the end of their tether so far as explanation is concerned. Some broader theoretical outlook is required if the treatment of these variables themselves is to acquire a deeper and more useful meaning. It is to the meeting of this theoretical need that the present article is addressed.

A word of admonition, however, should be said about its content. It must deal with issues that transcend psychology and pertain to all the sciences; for the problem of the nonquantitatively lawful in nature is universal. At first it might seem that the theory to be proposed, since it must be stated in correspondingly general terms, lies outside the scope of psychology proper. This impression would be erroneous. Though the reader may miss some of the familiar terminology, and though for want of space it will be necessary to ask him to make some of the applications for

[1] The writer gratefully acknowledges a generous grant of time by the authorities of Syracuse University for pursuing the theoretical and experimental studies upon which this article is based.

himself, the projected general model pertains at every turn to the task of explaining the fundamental processes of behavior. Among the problems upon which it specifically bears, within the limits of the present article, are the nature of psychological organization, motivation, learning, perception, and their interrelationship, the continuity-versus-discontinuity controversy, facilitation and inhibition, and the energies of attitudes. The writer believes that the theory is the *more* significant for psychology precisely *because* it has been developed in a wider frame of reference to meet a more universal challenge in science.

Let us begin by re-examining the role of quantitative statements in generalizations concerning behavior. The reader will find below a description of a familiar act or act sequence. The description is given wholly in terms of quantitative laws. Some of them are repeated as called for in the act, and nearly all are of the covariation type. The list is divided into five phases (A to E) to correspond to successive phases of the act sequence. Let us see how well the list describes and explains the phenomenon and whether we can identify the behavior involved.

A

1. The rate of evaporation varies with the temperature.

B

2. The curvature of the lens varies inversely with distance from the object of vision.
3. A neural impulse occurs at full intensity or not at all.
4. The magnitude of a neural impulse varies directly with the diameter of the neuron.
5. The intensity of a sensation increases directly with constant relative increments of the stimulus.
6. The more continuous the contour of an object the more readily it is perceived.

C

(3.) A neural-transmission impulse occurs at full intensity or not at all.

(4.) The magnitude of a neural impulse varies directly with the diameter of the neuron.

7. The energy of a muscle contraction varies directly with the number of muscle fibers excited.

Laws, 3, 4, and 7 now reappear in several repetitions, and, in connection with them, laws such as 2, 5, and 6.

D

8. The terminal velocity of a falling body is equal to the constant acceleration of gravity times the duration of the fall.

E

(3.) A neural-transmission impulse occurs at full intensity or not at all.

(4.) The magnitude of a neural impulse varies directly with the diameter of the neuron.

(7.) The energy of a muscle contraction varies directly with the number of fibers excited.

There are repetitions of this series, interspersed with earlier laws.

(8.) The terminal velocity of falling is equal to the constant acceleration of gravity times the duration of the fall.

9. The velocity of flowing varies inversely as the cross section of the flow.

If the reader now tries to state the behavior represented by the above list, he will probably be somewhat bewildered. He will guess that the phenomenon involves principles of gravitation and hydrodynamics along with behavior; but he cannot go much further than this with certainty. One wonders whether quantitative laws can be expected to *explain* an act sequence that they do not *describe* with sufficient completeness to permit its identification. Furthermore, there are important questions about the laws themselves that are unanswerable from the list: (*a*) Why are these *particular* laws brought together here, rather than countless others that could be mentioned? (*b*) How is their *order* in the list to be accounted for, including the repetitions indicated? (*c*) There appears to be very little "organization" of the laws in the list. Some organizing principle is needed to help us understand what is happening. We look in vain for answers to these questions in the laws themselves. There is something essential that they fail to give.

Let us now redescribe the act sequence in other, more familiar, terms, using the same interspersed capital letters to indicate the various stages:

(A) A small boy on a warm day, becoming thirsty,
(B) sees a pitcher of lemonade and a glass on a buffet.
(C) He goes to the dining room table, pulls out a chair, places it near the buffet, climbs upon it, pours lemonade
(D) from the pitcher into the glass,
(E) raises the glass . . . and drinks.

Now the matter is clear. By the use in this description of some terms *other than* quantitative we have been able to state the act sequence intelligibly. We are also able to answer, even on the basis of the meager knowledge of the events thus provided, all three of the questions *about the quantitative laws themselves* that these laws failed to answer. If we read the laws again, this time in conjunction with the acts listed above for the appropriate parts of the sequence, we can see the true basis of (*a*) the selection, (*b*) the ordering, and (*c*) the organization of the laws as they are "called into play." Something *other than quantities* of happenings has now been added to the picture, with a resulting increase of our understanding. The significance of the laws is clarified by showing that they are "contained," as it were, and ordered within a given self-delimited arrangement of happenings. Let us call this new "something" that is added to the quantitative laws in the form of a pattern of happenings the *structure* of the phenomenon in question. It will be seen that by this term we are *not* referring to anything that is "static." It is, rather, a *dynamic* structure—a structure of *events*.

Structure then, as thus defined, is what the quantitative laws fail to give; and it is what is needed (and needs more fully to be explained) if we are to have an adequate understanding of behavior. The structuring of events, to be sure, never occurs without the fact that the laws also hold good. Quantitative laws are demonstrable in all phenomena and are highly important and useful. Nevertheless our analysis shows that structure also is something that "holds good" and that must be considered in its own right. It seems possible that it may have laws of its own. If so, these laws will probably be of a different sort from the other (quantitative) laws; for we have seen that the latter do not describe the structure of phenomena and that the gaining of an inkling of the structure was necessary in order to answer certain questions about the quantitative laws themselves. We can go even further and say that if it were not for such a structure (comprising events of stimulus impingement, neural excitation, muscle-fiber activation, and the like), there would be no way of showing that the quantitative laws of behavior *exist*. The understanding of dynamic structure is therefore a matter of considerable importance.

Our statement of the laws in the example given was admittedly crude. Neither the definite equations nor specific quantities were given. If they had been stated more precisely and exemplified by quantities, would they then have been able to supply the necessary information and to identify the act sequence? It seems doubtful, since the added elements would still be abstractions so far as the actual pattern of events is concerned. It is true, also, that only a "sample" of the possible covariation laws was given. Suppose the list had been extended until it covered all the equations that could apply to a boy's getting a drink of lemonade (no doubt a very great number). Could the problem of the pattern have been solved then? This, too, is doubtful. If the laws were rigorously limited to quantitative statements, the task might have been even more difficult than before. We shall probably have to conclude, then, that the failure of the quantitative laws to be fully descriptive and explanatory does not lie in the paucity of the laws available nor in a lack of precision in their statement. It lies, rather, in their inherent limitation with respect to the problem in hand.

It is true that the glimpse of structure that illuminated the second approach to our example had to be inferred from a molar account, which, like molar statements in general, provided very little of the actual detail of structurization. Still, it served to recall, from other experiences, what we did know about the structure of the organism's behavior, including its neurophysiological aspects in their relevance to the episode in question. The most pressing present problem for psychology, in the writer's opinion, is to pass from such crude molar descriptions to a closer analysis and delineation of the *structure* of behavioral acts. What is latent or implicit *behind* molar formulations that can give the illumination which quantitative laws fail to provide? We should not beg the question by saying that unless structure can be stated in quantitative terms it is unlawful and cannot be explained. The quantitative laws *themselves,* even "molar" laws, require some structural understanding. Nor can it be said that structure lacks generality. What could be more universal in behavior than the "general format" of eating or drinking, or of a hundred other behaviors sufficiently stable and recurrent to have been given a name? But the problem is broader, even, than the field of psychology. Evidences of structures of a generalized sort occur in the phenomena of every science. If such structurings are general, and if they cannot be explained by quantitative laws, is it not logical to suppose that there may be such things as *structural* laws, that structure is something *sui generis?* There might even be some one universal structural principle that operates throughout the whole of nature.

The history of psychological schools and theories could, in a sense, be regarded as the record of attempts to deal with this problem in the field of psychology. They range from totalistic concepts such as gestalten, sign gestalten, supersummative wholes, cognitive maps, topological and brain fields, and "hypotheses," through open systems, mechanics of redintegration, communication and information, and mathematical brain models, to the stripped intervening constructs of behavior theory. Metaphysical postulates, also, such as "emergence," "entelechy," and various "manikin" assumptions have not been wanting. No final and conclusive answer has yet appeared. The very diversity of these efforts attests the difficulty of finding some clear, denotational way in which the structure of behavior can be described.

Perhaps the most general reason for the failure to solve the problem lies in the fact that it has not been approached in its own right. It has been assumed, in effect, that there *is no general* structural problem, that the means are already at hand for explaining each specific structure by the traditional methods of scientific logic. On the one hand it is assumed that, since every event must have some cause, the ordinary *logic of causality* should be able to explain the "structuring" of events. That we have not been able to explain matters this way is due merely to the fact that we have not yet found the right cause. On the other hand, there is a tendency to believe that the *laws* which state covariations and thresholds of measurable quantities should be able to bring it about that each element of a phenomenon gets placed in the proper spatial position at the exact time and sequence required for its characteristic structuring. The fact that no one has been able to show how quantitative laws accomplish this feat has not been sufficiently taken to heart. It is forgotten that quantitative laws are merely descriptive statements, not causal agents or forces. Frequently a "mechanism" of some sort (a term borrowed without justification from mechanics) is "postulated," through which the laws are believed to "act," or within which they are said to be "manifested," and into which "intervening variables" can be injected to fill the gaps in our structural apprehension. The writer has discussed the assumptions underlying the belief that quantitative, or mechanical, laws are the "architects of structure" in a forthcoming work (1). We shall here turn our attention to the problem of "structural causality."

II. CAN STRUCTURE BE EXPLAINED BY "CAUSE AND EFFECT"?

According to the commonly accepted definition of cause and effect an event, *O*, is the "cause" of an event, *P*, if it precedes *P* and is a necessary and sufficient condition of *P*. If *O* then *P;* if not *O* then not *P*. If we employ the symbol $\supset$ to indicate contingency and a superimposed dot to show negation, this can be expressed as

$$\begin{array}{c} O \supset P \\ \dot{O} \supset \dot{P}. \end{array}$$

Causes and effects can also be written as a linear series of single elements in which each succeeding effect becomes a cause for the next effect, thus:

$$\begin{array}{ccc} & M \supset N \supset O \supset P & \\ \text{---etc.}\rightarrow & & \text{---etc.}\rightarrow \\ & \dot{M} \supset \dot{N} \supset \dot{O} \supset \dot{P} & \\ & \underset{\text{time}}{\xrightarrow{\hspace{3cm}}} & \end{array}$$

Time is here represented as a linear stream, and no limit can be placed upon the number of cause-effect pairs that precede the series shown or that follow it. In psychological theories such *limited* series as shown above are illustrated by linear sequences such as stimulus—receptor excitation—afferent neural excitation—central neural excitation—efferent neural excitation—muscle-fiber contraction. Something like this is implied in the Hullian system in the "linkage" between the stimulus process and the reaction. There is, however, no logical reason why the causal series should not be extended indefinitely at both ends. We shall presently consider this possibility in the example of the boy who is getting a drink of lemonade. This historical method of causality explanation is seldom carried through in explaining the phenomena of behavior, or indeed anywhere else. A certain part of the chain is delimited as "belonging" to a particular behavioral act or other phenomenon; and the remainder of the sequence is ignored. The fact that this can be done without being aware of any arbitrariness is itself an evidence that some principle other than linear causation must be at work. To find this principle is our present task.

But it is probable that events do not happen in the single-chain fashion just indicated. *O* may be a necessary condition of *P*, but it is usually not a *sufficient* condition. Other events, *O′*, *O″*, *O‴*, and so on, must be present together with *O* in order to predict the occurrence of *P*, even though the absence of any one of them might negate *P*. And similarly, each of these *O*'s may have behind it another compound set of earlier "causes." This situation, which is still linear in its sequences, is familiarly known as "multiple causation." It is represented in behavioral theories by such concepts as stimulus compounds, drive stimulus added to object stimulus, and attendant reinforcing conditions. Such a compounding of causes makes the definition of causation as a necessary and *sufficient* condition somewhat inapplicable right at the start.

But to proceed further, let us note that a number of manipulanda in the environment are frequently required for the description of a behavioral act. The chair, pitcher, and glass were prerequisites in the boy-lemonade example. Let us call the events of contact with such objects *P, Q, R,* etc. Now behind each of these objects there lies a cause-effect sequence that extends indefinitely into the past and ramifies in space. Let us take the boy's contact with the chair as *P*. An earlier *O* existed in the form of someone's placing the chair at the table. Behind this, at a still earlier time, was the matter of purchasing the chair, a happening which might have had multiple causation (*O*'s) in acts of conferring between the boy's father and mother and in the combined acts of sales clerk and cashier. Behind these was the act of the store's manager in having previously "stocked" the chair, and behind this were many happenings involved in transportation, each of which, in turn, had its multiple precursors. For example, prior to the "loading of a chair" there were acts of a number of workmen in a factory handling tools and materials in the making of the chair. And again, prior to *both* the materials and the tools lay multiple occurrences such as wood cutting, fabrication, transportation, and so on. As the linear series is traced backward in time it is seen to spread out in space as a "regressus pyramid." By the time we have traced it backward only a few steps the number of *O*'s (and *P*'s) required to account for the contact of the boy's hand with the chair is so great that prediction from any one event has only a negligible value. Merely to illustrate the well-known regressus expansion (rather than for any intrinsic value it may have) Fig. 1

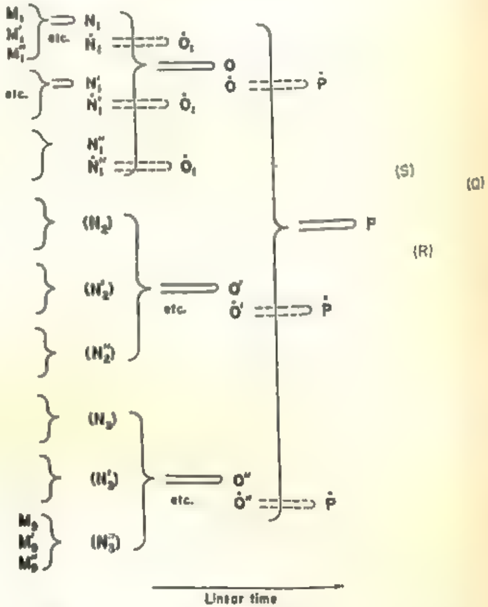

FIG. 1. Scheme showing the multiplicity of earlier events contributing to a final event *P* in linear causality. The letters in each column, with their braces, going backward in time from *P*, show three successive stages of the regressus. The symbol ⊃ indicates positive prediction (e.g., if *O* occurs then *P* will occur). A dot over a letter signifies *non*-occurrence of an event ("negative" prediction is involved). For illustrative purposes it is assumed that behind each event lies a "compound" of three other (necessary) events, no one of which, by itself, however, is sufficient for the prediction of the event in question. Thus no single event is both a necessary and a sufficient basis of prediction, or condition, of any other event. The number of contributing events at any given stage is equal to C^r, where *C* is the number of events in the compound (assuming that number to be constant) and *r* is the regressus stage taken.

is presented. The same analysis, of course, could be made with respect to the other objects, pitcher and glass, contacts with which were also a necessary part of the situation (*Q*, *R*, and *S* in Fig. 1). But now another difficulty arises. Each of these objects will have had its background pyramid series; and in order to produce the structure of behavior which we are considering the apices of these series must converge toward the interior of this particular dining room at a particular time when the boy also is there. How can we explain this convergence? We look in vain among the total antecedent events of the chair, the pitcher, the glass, and the boy, for any earmarks that will indicate a "destiny" of their coming together with the others at this time and place.

It might be argued that the boy's organism itself will supply the combining clue. The sequence is spelled out for us in the drive-and-stimulus-to-reaction chain of the mechanistic theories. But here again we shall meet with disappointment so long as we stick to the linear meaning of causality. First, we have the historical and environmental problem all over again in accounting for the *O*'s that lie in the boy's neural and muscular metabolism. But there is something more than that. There is built into the boy himself a kind of causal regressus pyramid. Neurologists have pointed out that as one proceeds (linearly) from receptor processes through central connections to effector units, a marked shrinkage in the number of elements or available pathways occurs. Many *O*'s are required for each ensuing *P* at every stage; and as one moves "backward" through these increasing sets of "multiple causes" the same sort of spreading effect is seen in the physiological regions as was noted in accounting for contacts with environmental objects. In terms of cybernetics it is said that the loss of information merely in proceeding from neuron to muscle is 100 to 1 (3). And again, we have no way of describing by any causal series within the organism how the various elements or processes are integrated in the pattern of a single act. The structure simply "appears" among the ongoing elements as though it were not "caused" at all.

One further lesson can be gained from all this. We find that when we try to explain structure by temporal trains of causes and effects, we are usually faced by structure as an accomplished fact. The practically simultaneous excitations that existed in the boy's sensory cortex as he surveyed the possibilities of getting a drink of lemonade were all there together. The coordinations of separate movements by which he gained his end were also contemporaneous arrangements. Then too, as we looked back at the historical sequences outside the organism, the acts of human beings or machines by which the chair was transported, manufactured, and so on, these also were seen to be matters of spatiotemporal patterning. The behavior of one individual in any one of those aggregates was coordinated with and dependent upon the concurrent behavior of others. Patterns seem to flow not from linear trains of causes and effects, but somehow from patterns already existing. Structures come from structures; and in many cases the structures themselves, as wholes, seem to operate not sequentially but in a contemporaneous or concurrent fashion. Unless causality is already set in this "framework of structure" it becomes merely a pyramiding manifold of happenings without relevance to the structural problem. But when it *is* placed in such a setting is anything of importance added to the picture by the notion of causality?

III. TOWARD A GENERAL THEORY OF EVENT-STRUCTURE

It appears, therefore, that the attempt to explain structure through the customary time series of cause and effect is futile. Some other explanatory concept must be sought that will circumvent regressus and link up events in some kind of pattern. Explanations must lie in the approximate "here and now" rather than in the remote past. The only way to accomplish this seems to be to cut across the conventional and absolute "time stream." One can think of time as the duration occupied by the successive ongoing processes and events of a particular pattern *that closes itself through a cycle of operation.* Taking this idea as a clue, we shall begin the presentation of our proposed theory of structure by stating the following postulate: *All structures of events have a self-closing or cyclical character.* Instead of depicting the events of any aggregate, *M, N, O, P,* or *P, Q, R, S,* as a linear series, we shall always try to think of their occurrence as shown in Fig. 2. If *P* starts the series, the event succession returns to *P,* or at least to the region in which *P* occurs. If it returns to *P,* it may, thereafter, keep on going in the same manner through repetitions of the cycle. If the cycle thus repeats itself, that would represent another "round" of "structural time." Time is thus always of the structure.

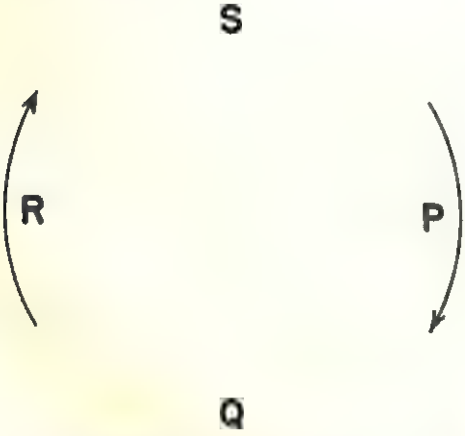

FIG. 2. Hypothesized arrangement of events in a (self-closing) structure. Arrows indicate event succession in "structural" time.

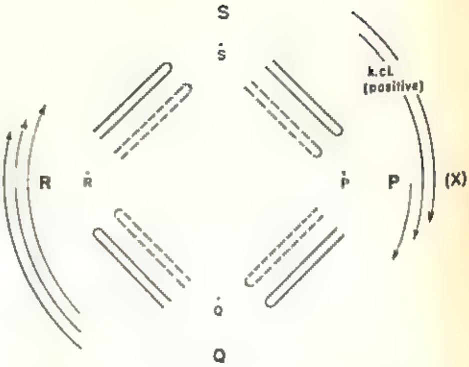

FIG. 3. Causality seen as dependent upon a prior hypothesis of structure. *P* is taken as the "starting" event. The cycle of events, as indicated by the arrows, repeats itself. *k.cl.* kinematic closure. It is maintained that only in such a predictively self-closing arrangement (structure) does "true causality" (i.e., positive *and* negative prediction by *single* event roles) occur. Contrast this figure with the compounding linear regressus of Fig. 1.

Though many such cycles of events may, of course, be connected by common events, regressus will be eliminated, first, because each structure preserves indefinitely its own characteristic pattern of ongoings, and secondly, because, though cycles can be linked indefinitely through space, their (repeating) operations may be actually "simultaneous" (that is, contemporaneous) in the linear meaning of time. The causality definition can here be reintroduced, if we wish, but this time in a structural rather than a linear setting. Figure 3 will illustrate this usage. The arrows suggest that this is a "repeating" cycle. We shall speak of the fact that the series returns to *P* (its starting point) as "kinematic closure" of the cycle (*k. cl.* in Fig. 3). In this case the closure is "positive" since it implies a continuation of the cycle. It should be recognized, however, that the reintroduction of causality, as shown by the symbols, is merely a device for semantic convenience. It does not contribute anything beyond the postulate of structure upon which it is here already predicated. *P* is the sufficient and necessary condition of *Q*, *Q* of *R*, and so on, only because of the postulated self-closing character of the pattern.

But in order to fit the facts of behavior there must be introduced a second logical construction, having a type of closure prediction different from the first. It contains, at the last event position of the cycle, a new type of causality statement that is the inverse of the old; namely, if *S* occurs, *P* (the starting event) will *not recur*, and if *S* does not occur *P* will keep recurring (tangency of the structure at *P* to some "outside" structure, *X*, is presupposed in the latter case). But again we note that it is the structural hypothesis that is fundamental, not the symbols or concepts of causality. This construction is shown in Fig. 4, and kinematic closure is here said to be negative. Such an arrangement typifies those cases in which the events of the cycle terminate upon a return to the initial region (nonrepeating cycle). It may follow, in some cases, upon a series such as that represented in Fig. 3. An illustration of the latter condition would be found in food-taking behavior (by taking repeated mouthfuls) as the "hunger contractions," *P* (perhaps from bloodstream events), continue in the stomach. After a certain number of mouthfuls are taken, the situation represented in Fig. 4 would occur, brought about, perhaps, by a tangent cycle involving "nutrient" events in the blood stream. Another example of negative closure is to be found in the breaking of contact (event *S*) with a hot object, an event through which the initial event of stimulation (*P*) is prevented from continuing or recurring. If the reader wishes to generalize these schemes of positive and negative kinematic closure, he will find that they have a very broad application to behavior. It might, perhaps, be objected that if we do not come back to an actual reoccurrence of event *P* (Fig. 4), we do not have a true closed cycle. We do, however, have a closed cycle in a fundamental sense, since, in order for *P not* to recur, there must be a change or "deflection" of some sort *in the region of P* that negates *P*'s recurrence. For example, in the breaking of the contact of the hand with the hot object, the events do bring us back to a change in the state of affairs in the *initial region.*

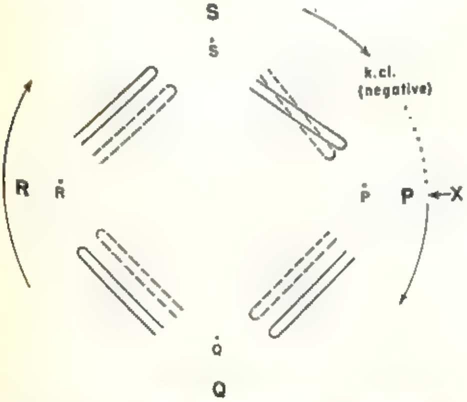

FIG. 4. Causality seen as dependent upon structure, with the predictive roles of *S* and $\dot{S}$ reversed (as compared with Fig. 3). If *S* occurs, the starting event, *P*, does not recur; the cycle ends in the "region" of *P*. If *S* does *not* occur, *P*, assuming that it is "fed" by an outside event, *X*, keeps recurring even though the cycle as a whole does not close. The cycle is thus one of a nonrepeating type.

We need one more addition to this purely logical stage of the model. Structures, wherever we take them in nature, probably never exist "in a vacuum." There is always "tangency" with other structures somewhere; and we can be sure that our structure $\begin{matrix} & S & \\ R & & P \\ & Q & \end{matrix}$

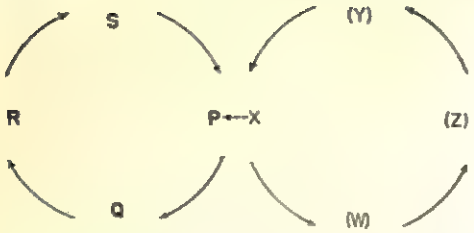

FIG. 5. A hypothetical event cycle (left) whose events (energies) are being augmented (or decreased) by interstructurance with a "tangent" structure shown at the right.

is not only operating at its own "proper" level or frequency of events, but is capable of receiving events (i.e., energies) from adjacent structures, or, perhaps, of losing energy to them by some kind of pre-empting of its elements. Let us represent such tangencies, with application to both the earlier nonrepeating and repeating models, by the symbolization of Fig. 5. *X*, which is an event (event role) of another structure, is not here regarded either as a sufficient or a necessary condition of *P*, in the sense of *P* as an event role. The position *P* would be the site of *some* events without the aid of the tangent structure. *X* provides merely an "energic reinforcement" of (or perhaps detraction from) the events occurring in structure *PQRS*. Linear causal regressus, in Fig. 2 through 5, has entirely disappeared. Aside from the structure *PQRS*, taken together with other structures (such as *WXYZ*) that may be immediately tangent to it, we have no interest in the nexus of events either in the past or the future. We do not care what the sources may be from which the contributing structure has, in turn, had its own energies supplemented, so long as the energies and the energic contribution of that structure to the main structure can be determined.

So much for the general format of the logical model. It was intimated above that the letters of the model are really event *roles*. The term "energies" was also used. These matters must now be explained. One of the defects of the notion of cause and effect, in addition to those mentioned, is that it implies that there is a specific event, *O*, invariably preceding and followed by another specific event, *P*. Empirical observation of event series, however, shows that such identifiable sequences of specific happenings do not invariably occur. We can say only that there is a certain *probability* that *P* will follow, and be preceded by, *O*. In fact, the occurrence of *O* (or of *P*) itself is a matter of probability dependent upon certain conditions. Instead of saying "if *O* then *P*," we might better say "probability of *O*, probability of *P*," and then add a third probability to express their *joint* occurrence or succession. Probability considerations, however, always imply that we must have a *fairly large number* of cases to observe. In order to determine probability or expectancy *many* events (or failures of events) and many successions or failures of successions must be counted under the classifications *O* and *P* and their interconnection. A further imperative reason for this pluralizing of the event concept within a role is the fact that we must make our model of structure general. It must fit all levels of nature (not just the macroscopic or molar level) or it will probably not fit any.

We might as well resign ourselves, then, to the necessity of treating events at the level of the microcosm (i.e., the most minute elements or happenings in nature) right at the start. And here, what has just been said of the indeterminacy of specific events and their succession applies with great force. There is, for example, no way of predicting that a certain minute particle, *a*, will collide with a certain particle, *b*, and hence there would be no prediction of a one-to-one series or closed sequence of such specific events. The best we can say is that for any particle there is a certain probability, through time, that it will be in a specified region and hence that it will be "available" there for an event of encounter with another particle. Whether one of these ultramicroscopic events will take place between two *particular* particles is "upon the lap of the gods." There is, however, a certain "probable density," or *probable number* of such encounters, that can *in the aggregate* be predicted to take place in a given region through a given time. If this is so, there is also a degree of probability that such (probable) encounters will occur in *all* the regions around the (hypothetical) structure. Cause and effect, in the usual sense, must therefore be replaced by a statistical treatment of the matter. Where the number of minute events recurring in a certain region is very large, so that they can be observed "macroscopically" or, as it were *en masse*, we say that "*the* event *O*" (at our level of observation) occurs. And if one of these (macroscopic) "events" regularly follows another in successive spatiotemporal regions of observation, we say "If *O* then *P*"–or, "*O* is the cause of *P*." This, however, is only a crude statement and one not at all suited for the careful study and explanation of structure.

The formulation of events in terms of probability holds, of course, for all the event positions labeled *P, Q, R, S,* and *X* of the preceding diagrams. (In Fig. 4 the probability in position *P*, following the occurrence of events at *S*, falls abruptly toward zero.) In order to apply these concepts it is evident that we must always regard *P, Q, R,* etc. in our logical model not as single definite happenings, but as indicating *regions of space through time* in which events may or may not occur. They are the "event regions" that are hypothesized as defining the structure. We must also be prepared to conceive the events in vast numbers in any region, and in ultramicroscopic as well as in macroscopic terms. But let us remember the further aspect of probability that must be incorporated in the design. Having dealt with the probabilities that events will occur in each of the regions *P, Q, R,* etc., singly, we now have to consider the probability that they will occur (with sufficient probable density) in *all* these regions of the self-closed structure taken together, that is, around the cycle, either simultaneously (if continuous) or in immediate temporal succession. As this probability approaches 1.0, it means that the structure in question is becoming increasingly clear and predictable. It is hypothesized that this is the phenomenon that takes place in both learning and perception.

Our *second* postulate, then, is as follows: *The observed occurrence of a structure of happenings is dependent upon* (a) *the probability of occurrence of events in each of the* (*event*) *regions of the structure, the regions being taken singly, and* (b) *the probability of the joint* (*or successive*) *occurrence of events in all the regions of the struc-*

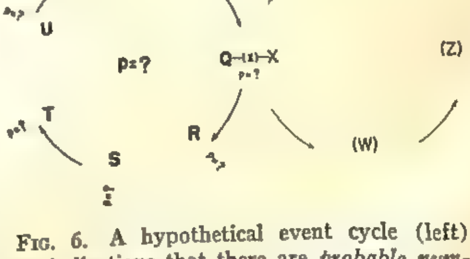

FIG. 6. A hypothetical event cycle (left) with indications that there are *probable numbers* of events (event densities) in the respective "event roles." Capital letters now signify space and time *event regions*, rather than single events. An interstructurant cycle is again shown at the right.

ture. This postulate is symbolized in Fig. 6 in which a larger number of regions is employed, the main structure under consideration being shown at the left. In this figure *p* indicates the probabilities within the single event regions represented by capital letters, and *p* the probability of the structure as a whole. As before, a contributing tangent structure is included for completeness. The *I* between *Q* and *X* represents an expected "interstructurance ratio" between increases of the events in the regions of structure *WXYZ* and those of the main structure. If the main structure is a food-taking behavior cycle, the contributing structure might be a cycle of events in the blood stream. Solid arrows indicate the temporal succession of events in the regions of the cycle.

Before going further with the model, which is still, for the most part, only in a logical stage, let us make some direct applications to the organism. Again we shall discuss the example of the boy getting a drink of lemonade, but shall ignore the contributory cycle (at the right in Fig. 6) and consider only the main structure. Figure 7 presents, for this purpose, a cycle having a greater number of event regions. Probability symbols are omitted *but should always be understood,* both for the event regions and for the structure as a whole. Let us first establish a clearer definition of the term "event." This we shall define solely in terms of an indivisible, all-or-none, happening, as, for example, in an encounter or collision where minute particle elements come together (or, in relativity theory, to "near points") and then go apart again as they continue on their courses. Events may also be the sudden breaking of contact between elements. The dichotomous states involved in ionization and chemical interchange could come under this general definition of events. Such a definition may not seem at first to fit happenings like action across synapses; but it is believed that, if conceived at a fine enough, ultramicroscopic, level, it can be considered appropriate. *A single event, then, is a "dichotomizing," nonquantifiable, happening, and nothing more.* Its representation on a spatiotemporal model would be merely a point. We shall sometimes, for convenience, speak of "an event" in the singular when we mean a large number of such events (that is, an event role); but it should always be remembered that the letters of the diagram (Fig. 7) represent event *regions* of space and time in each of which a *large number* of these elementary events occur, giving us, when they occur at a probable density of threshold frequency, "the event" as *macroscopically* observed. Thus, as we know, a large number of stimulation points are involved on the retina or the skin as the boy sees the pitcher and the chair or as he touches these objects. A large number of afferent-neuron excitations and cortical synaptic events are involved in the sequelae of these stimulations. A large number of muscle-fiber-activation events occur at the efferent end plates. Many molecules of liquid strike the boy's throat as he drinks, replacing the many that have evaporated from the throat membrane. Each of these pluralities can, of course, be regarded as

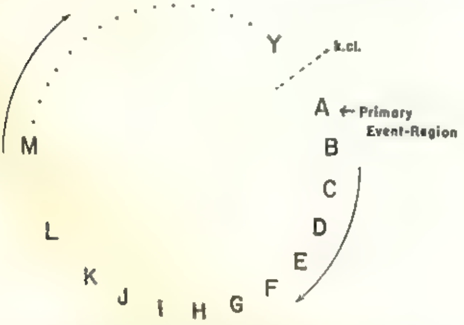

FIG. 7. A hypothetical event cycle of an act (see text)

the compounding of events in their more elementary status at the molecular or atomic level. It is useful to regard some one (or more) of the event regions of the model as a *"primary"* event region in that it represents the *initial* marked increase (or decrease) in events (energies) through tangencies with an outside structure. This procedure also locates the point of closure, which comes just before the primary event region (*k. cl.* in Fig. 7).

Let us now represent the event of "throat drying" or, more exactly, the region in which many microscopic events of tissue change occur as the body moisture evaporates, by *A* (Fig. 7). *A*, then, is the primary event region. Omitting vision and some other aspects for simplification, we can now assign to the other letters, as regions, approximately the following event roles: *B*, stimulation of receptor(s) in throat membrane (from the drying); *C*, excitation of afferent neuron(s); *D*, excitation of neuron(s) at synapses in the central nervous system; *E*, excitation of other neuron(s) at other synapses; *F*, excitation of efferent neuron(s); *G*, excitation of extensor arm muscle fibers at end plates (hand here moves forward toward the chair); *H*, contact of hand with chair; *I*, stimulation of proprioceptors and tactual receptors by this contact; *J, K, L, M*, etc., afferent, central, efferent, and muscle-fiber excitations (as hand *closes* on the back of the chair).

From this point on let us simplify matters by conceiving further elements of the series (dotted line) as representing other neural, synaptic, muscular, receptoral, and bodily contact events as the chair is placed in position and as the boy climbs up, takes the pitcher, pours a glass of lemonade, and tips the glass up at his (open) mouth. Eventually we come to an event (let us call it *Y*) at which the liquid encounters the throat membrane and the "moistening" of the throat represents a partial negation (diminution of energies) in the "drying" events of the tissue. With kinematic closure at *Y*, then, we come back to the starting (or primary-event) region. Since the tissue-drying events are only partially reduced in number, a repetition of a part of the cycle will occur (positive *k. cl.*). This part, not distinguished in the diagram, could be charted as a "component" cycle involved in the taking of *successive swallows*. Finally, as the "swallow cycles" continue, event densities at *A* are reduced to a state at which the whole cycle is energically "in equilibrium." Events at *A* now cease to occur (negative kinematic closure) and with their nonoccurrence the remainder of the cycle is negated, at least in so far as the superthreshold level of "conduction" and overt action is concerned. It should be noted (though it cannot be explained at this point) that some of the event regions in the cycle may be more "readied" than others. That is, they may already have an event density approaching threshold, though the remainder are not yet at a probability stage in which the structure as a whole can appear. *Continuation* of the occurrence of events in the "readied" regions (see later) would guarantee their presence when needed for the total structure. (Many considerations that would need to be included in a full structural diagram, such, for example, as a coordinated cycle of mouth opening and [later] closing, have been omitted to simplify the illustration. Tangent "feedback" cycles of optical and opposed neuromuscular ongoings have also been omitted.)

There is still one feature that must be added to the logical model before it can be given its full physical or organismic significance. The events *A, B, C*, etc. of Fig. 7 are really *connected*.

Is such connection only a matter of abstract joint probabilities, or is it "physical"? Theory and common sense both require the latter answer. If events represent encounters between minute (hypothetical) elements that collide or come to "near points" in their ongoings, then the only way in which such events can be connected is by the fact that one ongoing element, after it encounters its opposite, continues and makes an encounter with another ongoing element. To illustrate such ongoings in our example let us recall that water molecules *travel* in space as they "evaporate" from the throat membrane. Some kinetic feature is probably also present in the molecular activity of receptors, connecting stimulus event with afferent-neuron-excitation event. A neural impulse represents a *whole train* of *minute cyclical ongoings* (of ions) through and along the neural membrane, connecting the events of excitation at one end of the neuron with events of excitation at the other. The more grossly perceived ongoing of the hand as it raises the glass, represented at a microcosmic order by neural and muscular ongoing cycles, connects the event region of "glass grasping" with that of "glass tipping." The flowing (or fall) of the liquid is the descending portion of a gravitational cycle of ongoing and connects the events of displacement of the liquid as the glass is tipped with the events of the droplets striking the throat. In this way a connection of events is provided *by each ongoing role.* There is also a connection of the ongoings *by* events. There is, in other words, a structure of both ongoings and events. A kinetic or "motion" aspect must therefore be added to the elements of the model. To this task we shall presently return. Again we note, from the format of the model, that these several ongoings (which are in general themselves cyclical) and the events by which they are connected and which they connect form, when taken together, an *over-all* cycle (Fig. 7). And again, the *numbers* in which events occur in each of the regions separately, *and therefore the probability of the behavior structure as a whole,* are a matter of the (sub)microscopic probability density at the event regions.

It will be seen that regressus is completely eliminated from the model. All that is needed is that there be a sufficient space and time availability of ongoing elements at event regions necessary to constitute a structure. The pitcher of lemonade might have been on the buffet many hours, or it might have been there for only a thousandth of a second before the boy's eyes turned toward it or his hand encountered it; and it could have come there through any one of an indefinite number of pyramiding lines of "causality." These considerations are without significance for our present problem. The only thing that concerns us about the concentration of molecular cycles we call the pitcher and its contents is the *probability* that such a "concentration" will be present at a time and place that will permit encounters to be made with it by the "ongoing elements" of the boy's hand. Since, however, all events have a *certain degree* of randomness, this requirement makes room for "approximations." If, for example, the pitcher had been in another room, or had been available only just *before* the maximal drying (events) of the boy's throat occurred, the probability that a lemonade-getting-and-drinking structure would have occurred would have been *less* (though still not necessarily zero). And the same can be said for probable density at all the regions of the structure, including its tangencies with other structures, such as those of blood-stream events, which lie inside the organism.

It is unnecessary to ask what "makes"

or "brings about" all these regional event probabilities and their concurrence in time and space. They are implicit in the empirical situation itself. If they were not, the structure (that is, the behavior) would not occur. No "field expectancy," "organizer." or "entelechy" is needed in a theory of event structure. We can think of the pitcher of lemonade, the glass, the chair, the internal systemic changes that occur in "thirst," and all other relevant situational features as *"bounding* conditions" of the lemonade-taking-and-drinking structure. These bounding conditions, which are themselves self-closing *structures*, increase, by imposing space-time limits upon the freedom of adjacent ongoings, the probability that the (bounded) ongoings of the cycle under consideration will come to events at intervening regions (including synaptic areas) with a density above the macroscopic threshold; and in so doing they help to bring about the self-contained and self-closing structure of the (bounded) behavioral act. The implications of this conception for a new and more comprehensive theory of learning are evident. The notion of the probability density of a structure's occurrence under given bounding conditions might be substituted for such earlier notions as sign gestalten, "stimuli" linearly evoking "responses," strength of associative S–R linkages, and the selection and fixation of neural pathways. The difference between continuity and noncontinuity learning may be merely a difference in the shape of the curves of distribution of the increasing *probable structural densities* of the act as the experimental situation is repeated, the curve in each case being plotted along a continuum of structurizations or "trials" with experimentally imposed bounding conditions that are different in the two cases.

Let us now return to the problem of representing the motion of ongoing elements in the model. In order to supply this feature, the ongoings must be represented as *curves*, each suggesting *continuous* motion; for it seems probable that at the minute levels of nature particles *are* in continual motion and that their motion is cyclical or vibratory. Moreover, nothing ever starts from a position of ascertainable "absolute rest" or proceeds to another point of absolute rest. In fact, if we are to get away from a purely static conception, points in space and time are *definable only* as the points of conjunction between ongoings or motions. Let us try to diagram the situation, at first, without any consideration of where the ongoings start or of their ultimate destiny. Let us also avoid trying, for the moment, to link the model too closely to neurophysiological considerations. Figure 8 shows six ongoings (broken lines *i* to *v*, and *r*) with event regions between adjacent ongoings. Evidently we must assume that there is something in each case that "goes on." Without trying to be more specific, let us postulate, for the purpose of the geometry of the model, that it is, *in the last analysis*, a "continuance-head" (or particle?) of the smallest conceivable magnitude. (If the diagram were adapted to our present example, the *compounded*, or higher order, ongoing elements would represent such features as ion cycles in neural impulses, muscle-fiber molecular lengthenings and shortenings, flowing of the liquid, and so on.) The "event-connecting" segments of these ongoings occupy the central portion of Fig. 8. These ongoings must, of course, be conceived in *plural number* for each ongoing role; and we know, in fact, that this is actually the case. In any act of behavior, after many points are stimulated on a receptor surface, many neural impulses travel, as it were, "in parallel," many cortical fibers are involved, many

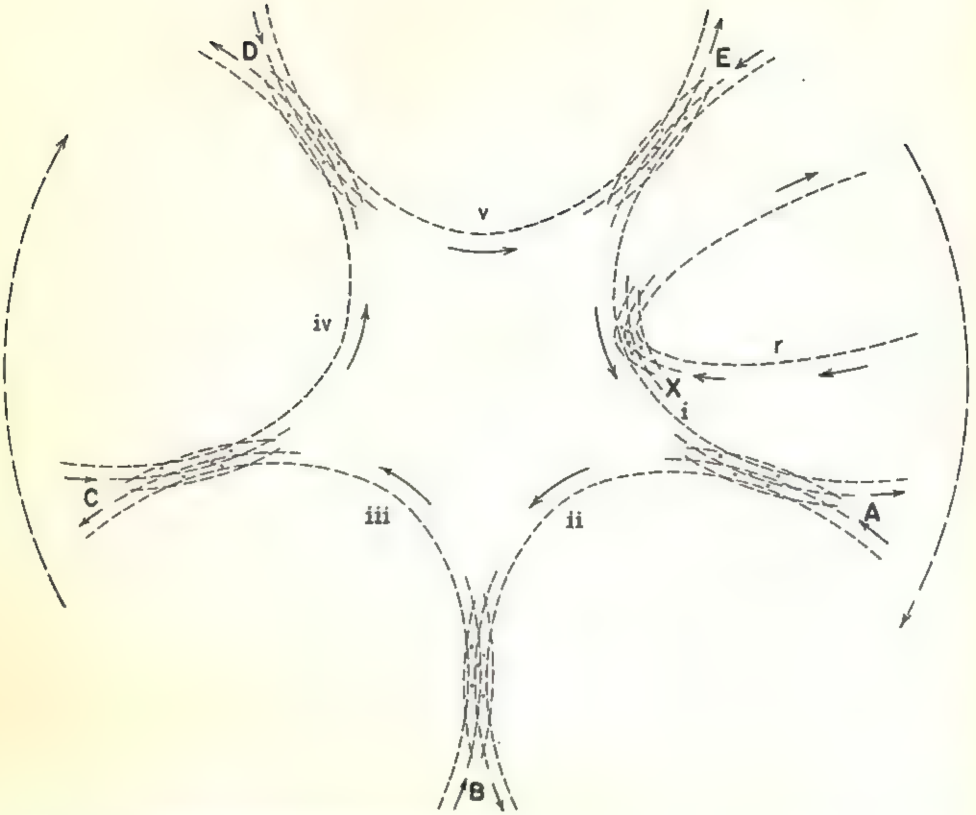

FIG. 8. The structure (?) of a behavioral act, with its ongoings connecting the event regions, as it would appear if the ongoings were conceived as indefinitely extended (i.e., linear). *i-v*. Ongoings (really pluralities or "sheafs" of ongoings, as shown at event regions). *A-E*. Event regions of the (assumed) structure. Dots in the regions represent events. *X* is an event region where an interstructurant ongoing, *r*, contributes events (energies) to the main structure. If the arrows in the central part of the figure *really* identified a temporal succession, the figure would symbolize a (self-closing) structuring of events; but it would not represent a completed structure of events and self-closing *ongoings*. Structure, in other words, is *not pervasive* in this model.

muscle fibers contract in their common role in a single effector movement, and so on. Note that this feature is suggested in Fig. 8 by the duplication of the lines for the ongoings (shown only in the areas of the event regions). Events between these multiplicities of ongoings, whose probable density or numbers in the several regions underlie the structural probability of the act, are indicated by dots. Ignoring ongoing *r* and region *X* for the present, we shall regard ongoing *i* and region *A* as our starting point. The short solid arrows indicate the direction or sense of the ongoings; and they show a temporal clockwise succession of event occurrences in regions *A* to *E* and back to *A*, as indicated by the longer broken arrows at the margin of the figure. We now have a cyclical structure of *events* (but of events only), connected by continuous ongoings which themselves are *not* structured but extend (small arrows at the periphery) from an indefinite past into an indefinite future. Let us see if this construction is satisfactory.

At least three provisions must be made for any structural model of behavior: (*a*) It must accommodate itself both to relatively stationary and to "successive" patternings. In some structures, as, for example, in perceptions of objects, all parts of the object seem to be perceived at once. On the other hand, many typical structures of behavior have a cyclical *succession* of happenings (as we have shown, for example, in the case of the boy and the lemonade). (*b*) In order to meet the latter requirement there must be some suitable arrangement for the timing of the ongoing elements so that the events (regions) will occur in proper order. (*c*) Many behaviors not only occur in sequential arrangements but are sustained, in the sense of a repetition (of their complete cycles), through time. This feature, which is called *steady state*, requires a continual input contribution of events (energies) from some explicable source.

An examination of Fig. 8 shows that the construction there presented does not meet these requirements. Time coincidence of ongoing elements in an event region and time *sequence* of the *successive* regions are shown but not explained; and in order to account for them we would probably have to invoke some special "organizing agency." "Static" or simultaneous event structuring could be accommodated if we consider that, instead of separate "volleys" of particles coming to events in the region, we have a *continuous flow* of particles along the course of each ongoing. Events would then be occurring in the five regions (*A* . . . *E*) in practical simultaneity. The situation so represented might be equivalent to an "equilibrium" of the structure. If we should wish to turn the picture into a kind of succession, we could do so by adding an "input source" in the form of another ongoing stream (*r* in Fig. 8) which comes to an event region (*X*) with one of the ongoing streams of the main structure. If the *increases* in numbers of events introduced into the cycle at *X* are passed on from one stream of ongoings to another, we would then have *successive* increases of density in regions *A* through *E*, and thereafter, under conditions of positive kinematic closure, around the cycle repeatedly. We can suppose that there will be some sort of output tangency to keep the state in balance, so that the *full* energies continually being added to *A* from *X* are not passed back from *E*, via i, to *A*, but only a portion of those energies. This interesting explanation, which may be called a theorem of "conduction," gives a basis for steady state. So far so good. But a difficulty arises in these explanations both of equilibrium and of steady state. Some continuous source for the ongoings (i.e., ongoing elements) is needed in both instances. Where, for example, does ongoing *r* get its supply? Either we must think of these sources as the ongoing lines themselves, extending from an infinite past, a conclusion at odds with the temporal self-containedness of phenomena, or else we must suppose that they are derived, for each of the unclosed ongoings in the figure, from tangencies with other, more remote, sets of (unclosed) ongoings. In the latter case we begin to slide back into the old multiple regressus.

The difficulties here encountered can be summarized by saying that we have been trying to build a self-closed structure out of materials that are themselves unstructured. One cannot make a true structure out of open-ended lines that merely "butt against" one another as in Fig. 8. Structure must be pervasive if it is to exist at all. It must be composed of units (in this case, ongoings) that are themselves self-closed. We need, therefore, to suppose that the six ongoing roles of Fig. 8, instead of extending out indefinitely in space through time, have a curvature throughout their course and return upon themselves. Could we say, perhaps, that the ongoings follow the curvature of the continuum of space-time?

Without trying to elaborate the last proposal we shall pass at once to a new and final design in accordance with the idea just expressed. It represents merely an extension of our first postulate (self-closedness) down to the lowest orders of the microcosm. For cartographic convenience only six subcycles of ongoings and six event regions will be used. Actually, of course, there would be a *very large number* since the model must be conceived, ultimately, in microcosmic terms. In Fig. 9 we have shown this structure, in principle, as 1, together with out-structural tangencies with two other structures, 2 and 2*a*, affording an input or added increment of event density, and an output, respectively. The legend will recall the meanings of the various symbols, and the earlier organismic details given for the boy-lemonade episode in connection with Fig. 7 (or any other behavior by which the reader might wish to test the model) will supply illustrative content. In applying the construction of Fig. 9 it should be remembered that there is no limit upon the number of subcycles of which the structure under consideration can be composed. We now have a consistent theoretical model of a structuring of events, in this case the structure of a behavioral act. It consists (see structure 1) of a set of subcycles of ongoings and a cycle of events (event regions) between, and provided by, the ongoings—a cycle of cyclical ongoings and events. Though the evidence cannot be here fully presented, but must rest with the specifications of the microcosm previously mentioned, the writer believes that the gross and finer facts of neurology, physiology, and environmental contacts will justify the use of such a model for the description of behavior. It is also consistent with structural principles to assume the existence of *"higher" orders* of structure, that is. of structures that are composed of cycles of cycles of cycles, and so on. For example, conceive a larger cycle made by joining a number of cycles such as 1 in Fig. 9. Structure can thus be pervasive and can provide an explanation of the various levels or "hierarchies" of nature. (As the reader will see, this was impossible with constructions like that of Fig. 8.) Such higher orders, for example, might describe the *collective* or *"social"* structurings of the behavior structures of individuals; and such a description could well replace the present ambiguous and unsatisfactory term "group."

The three requirements listed earlier for a structural model are all met by Fig. 9. Equilibrium or a "static" structural condition is achieved under certain conditions by the fact that the ongoings of each of the subcycles (see smallest arrows) return again after one event region to the region where they had been just previously. With a continuous flow of elements around each of the ongoing subcycles, a virtual simultaneity of events would occur throughout the structure.[2] But whenever a sudden increase in the probable density of events is given through a tangent input structure (2 in Fig. 9), the equilibrium of the structure is disturbed; and this increase, beginning with the primary event region, that is, the re-

[2] The vibrations of molecules in solid, colloidal, or liquid states might be an example of such an equilibrium. Postural-tonus is also suggested, though this may also have something of the character of a steady (repetitive) state of the whole cycle.

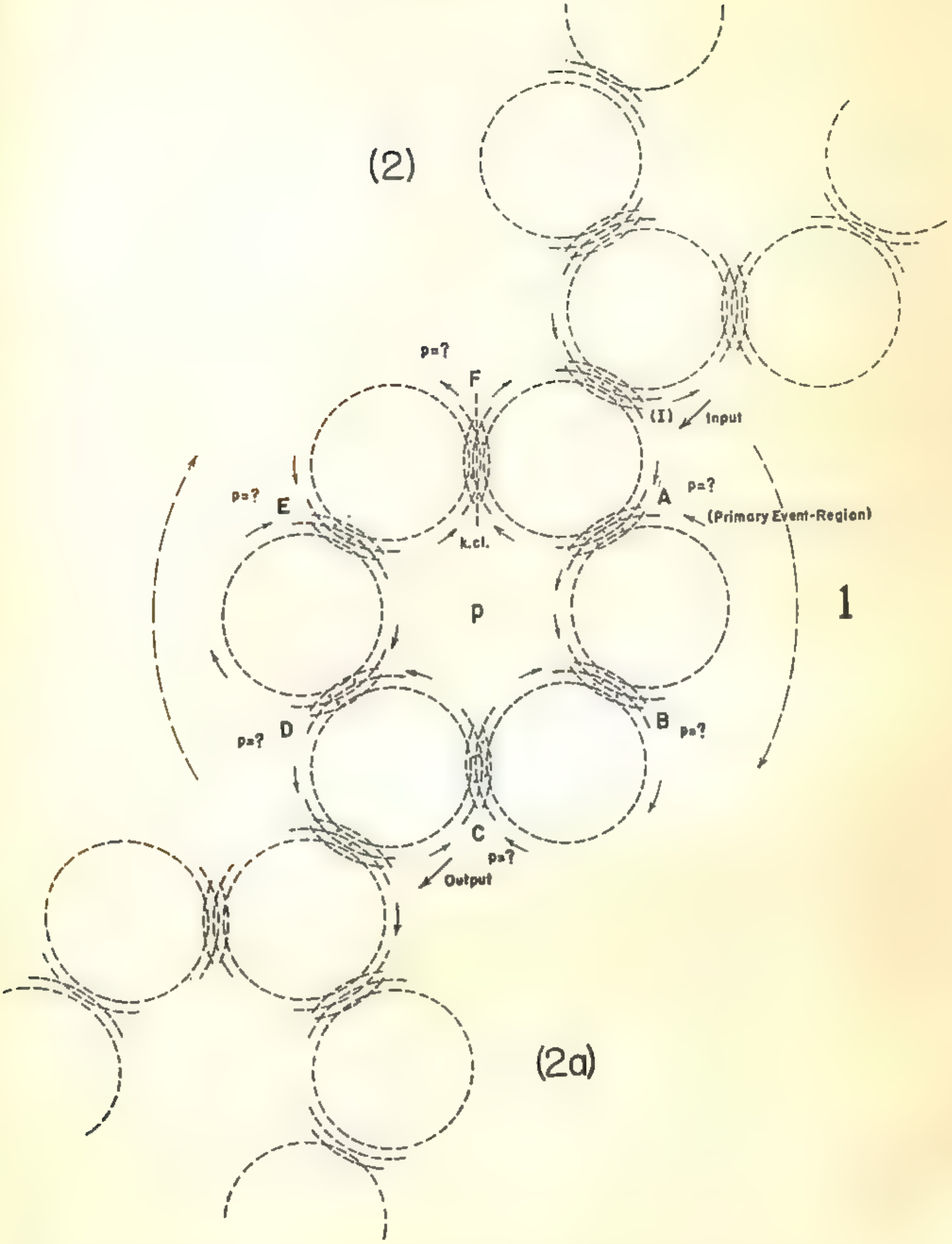

FIG. 9. Hypothetical diagram of the event structure of a behavioral act. 1. The event cycle of the act under consideration. 2, 2*a*. Tangent cycles providing "input" and "output," respectively. Broken line circles (see 1) are the self-closed ongoings of subcycles, with duplications to show plurality in each "role." Small arrows show direction of these ongoings. *A–F*. event regions of the act cycle. Dots indicate events (as of encounter) between the ongoings. *p*, *p*, etc. probable density (or number) of events (energies) in a region. *p*. probability of the occurrence of act structure as a whole ("structural probability"). *k.cl.* kinematic closure (+ or —). Broken line arrows show the order of succession of energic increments received from structure 2 as they are passed along through the event regions. *I* represents "interstructurance" of structures 1 and 2; it also refers to an "interstructurance index." (Ongoing and event details are completed for only a portion of the figure.)

gion first increased, is displaced around the cycle of event regions (broken arrows). Such a situation occurs in the full phasic sequences of a behavioral act and perhaps even at subthreshold density levels. With *repeated* rounds of the event cycle, should these occur, we would have a steady state. The problem of coincidence timing is solved by the reverberating character of the subcycles, which, with particle-elements continually ongoing, provide a *continual* availability for events. Succession in time could be guaranteed, in the sense of a succession of energic increases, by the passing along of increments from one event region to the next in the steady-state condition mentioned above. There is now no difficulty in providing for the *source* of the energic units of the cycle, or of the energic contributions made to the structure by adjacent structures. For we are not faced by a regressus to indefinite origins. The subcycles themselves, and the cycles of the tangent structures, continually possess availability for events in the immediate present by the self-closing and repetitive character of their ongoings. And we need to consider only the event cycle of the behavior in which we are interested plus the cycles that are immediately adjacent to it.

The task outlined at the beginning of this article has now been, at least in part, accomplished. It has been shown that by laying aside linear models and linear causality, and by passing over into a logic of structure supplemented by probabilities, we can arrive at a fairly clear conception of "another type" of natural law. The conception of such a law is essentially "geometric" rather than quantitative. Our statement of this (hypothesized) law, however, has been as objective in principle and precise in reference as statements of laws that are based on "abstracted" measured quantities. The writer maintains that the proposed paradigm is general for all acts of behavior and that it even suggests a unifying bridge across the "hierarchy of the sciences." Event-structure theory, though it includes quantitative considerations, nevertheless rests upon a foundation that is basically nonquantitative. Consider, for example, such concepts as the self-closedness of ongoing, the forms of kinematic closure, the indivisibility of an event, and the compounding of structural orders. No statements of dimensions, counts, or measurements can convey the full and essential meaning of these conceptions. The system is essentially a geometry of ongoings and their interrelations at "event points." In short, it is a theory of *structural kinematics*. But at the same time a *place* for quantities and covariational laws is provided. For quantities are probably related, in the last analysis, to energies; and energies can be represented, as we have shown, as *structured potentialities for numbers of events*. Hence the entire conception embraces also a theory of "*structural energics*" or "*structural dynamics*." It is important to note, however, that a rational place for quantitative laws could not have been provided without the (nonquantitative) structural kinematics.

The writer believes that the event-structural paradigm will be found to be applicable to all organismic phenomena at the biological and physiological, as well as at the behavioral, level, and that it will apply also to collective or social aggregates. Obviously the theory, as here proposed, is only in a pioneering stage. The account here given is also merely a general outline from which many details have had to be omitted. Many questions arise for which answers must later be provided; and a large amount of work, both experimental and theoretical, will be needed in order to arrive at a true appraisal of its validity. A further, though not a final, step in the theory's presentation, in which some of these questions are answered and additional properties and principles of structure appear, will be undertaken in a forthcoming volume (1). The theory will there be approached and exemplified through the facts and theories of perception.

IV. QUANTITATIVE ASPECTS OF EVENT-STRUCTURE THEORY (STRUCTURAL ENERGICS)

It is hoped that nothing that has been said will be construed as a failure to realize the importance of quantitative methods in the work of science. The quantitative and nonquantitative aspects of investigation should proceed together. An attempt has therefore been made to deduce some quantitative hypotheses from the over-all model, and, through the aid of students and associates, these hypotheses have been tested experimentally. Space will permit only the briefest description of this work.

The major hypothesis, thus far, has been concerned with predicting the amount of energy in a structure (cf., for illustrative discussion, structure 1 of Fig. 9). Such energy or probable event density is conceived under two aspects, the "autonomous" or "proper" energy of the structure, which is called the energy of *"structurance,"* and the energy that is contributed to it by other structures in "constructurance" with it, such, for example, as that shown at 2 in Fig. 9. There may be few of these or a great many, but all must be considered. Such contributions constitute the energy of *"interstructurance"* or the "interstructurance increments" provided to the main structure by its surrounding tangent structures, or *"manifold."* But the structures of the manifold may, in some cases, be *"anti*structurant," rather than constructurant, to the main structure. That is, there may be some kind of kinematic "deflection" which *deprives* the main structure of energies by decreasing its probable density, so that, instead of having the two structures increase in energies together when the tangent structure receives increases from *its* manifold, we may have a *decrease* in the energies of the main structure as the tangent structure is increasing. The contribution to the main structure from manifold structures (e.g., from structure 2 in Fig. 9) will, in such cases, have a minus sign. The constructurant relationship represents instances of *facilitation* in the biological, behavioral, and social realms, while the antistructurant relation represents *alternation, inhibition,* or *"conflict."* Furthermore, it is conceived that such additions or subtractions of energy proceed by constant (kinematically determined) ratios to increases in the manifold structure. This ratio (increase in the main structure divided by the attendant increase or loss in the manifold structure) is called the *"index of interstructurance"* of the manifold structure to the main structure. The interstructurance index (suggested as applying at *I* in Fig. 9) is specific to the pair of structures concerned, is limited to unity, and may be either positive or negative. (Actually it should represent a function expressed by a curve showing the relationship of the two variables.) The "output" quantity of the structure (for example, to *2a* in Fig. 9) can be neglected in this problem, since our objective is to find the amount of energy which is *available* in the main structure at a given (present) time, either for self-maintenance of the structure or for being passed on to adjacent structures. Hence the *total amount* of energies of a structure, so defined, at any given time, is a function of (*a*) its proper or "structurance" energies, which represent a sort of mean of its operation through time, or, if one prefers, a "homeostatic level," and (*b*) the sum of the interstructurance increments (or decrements) that are being received from its manifold. It is believed that these increments or decrements are given by the structurance energies of the manifold structures concerned, the latter, however, being first weighted, respectively, by their indices of interstructurance with the main structure. In the symbolism of the diagram (Fig. 9) this total amount of energies corresponds to the total number of dots in all the event regions of structure 1 (remembering that these have been augmented, or detracted from, through interstructurance with manifold structure 2 at I).

The generalized equation that has been derived in this manner from the postulates, kinematic concepts, and definitions of the theory is stated as follows:

$$E_1=f(S_1+S_2I_{2\to1}+S_3I_{3\to1}\cdots+S_nI_{n\to1}),$$

where the subscripts indicate different structures (1 being the main structure under consideration), E_1 is the total energy of the main structure, S is the "proper" energy or structurance value of a structure, and I is the index of interstructurance of a (manifold) structure to the main structure as shown by the subscripts and arrows. Structures 2 to n represent all the structures of the manifold; and the summation is, of course, algebraic. This equation is known as the structural-energics (or structural-dynamics) formula. In using the formula, manifold structures considered on logical grounds to be very low in S or in I are usually omitted as having no appreciable value in the summation. Should the value of the equation turn out to be negative, E_1 would not represent "negative" energies occurring in structure 1, but presumably (positive) energies that were being expended in a structure or structures *antistructurant* to 1. The manifold of structures (shown to the right of S_1 in the equation) can be broken down into various classes or types, and the contributions of these types, separately, to the dependent variable, E_1, can be determined. For example, in predicting the intensity with which an attitude is held by an individual (energics of an attitude structure), as E_1, three types of manifold structures have been used, viz., personality-trend structures of the subject, small "face-to-face" collective structures into which his own behavior is structured, and his larger organized, or institutional, structures. The total summation, as provided by the equation, is hypothesized as giving a fairly accurate prediction of strength for the attitude concerned.[8]

The testing of adapted forms of the equation has thus far been carried on by simple correlation procedures involving E_1 and the manifold summation only. Methods of measurement for S and I, and sometimes for E, have been limited to subjective scaling, but with carefully prepared forms. The actual "energies" implied in E and S have thus far had to be inferred from the reactions on response forms or in an experimental situation. The equation has been tested (or tested in part) in nine independent investigations, largely at the structural orders of personality and social psychology, and representing cases from the fields of propaganda, attitudes, personality characteristics, job adjustments in industry, insight, learn-

[8] Since the attitude represents a "meaning structure" within the individual, it is assumed that the two types of collective structures interstructurant with the attitude cycle are also represented at the intraorganismic level (i.e., as meaning cycles).

ing, and custom behaviors of American males. In all but one of these investigations (an early one in which it now seems that the hypothesis was not adequately stated) significant correlations, ranging approximately from .20 to .80, were obtained. Worthy of note here are the variety of structures whose total energies were to be determined, the considerable number of independent raters often employed for determining the different variables, the very large number of manifold structures whose increments entered into the summation for each subject in most of the experiments, and the fact that these interstructurance increments were either positive or negative and hence involved subtraction as well as addition of energies in computing the predicted *E*. The consistent experimental support given the hypothesis in the face of these complex conditions seems surprising and would tend to suggest that, although it was developed in a highly general frame of reference, or perhaps *because* it was so developed, the theory does accommodate itself to the quantitative facts of human behavior.[4]

V. Summary

The adequacy of quantitative laws and the common practice of thinking in terms of linear cause and effect, as methods of dealing with the universal problem of structure, are questioned. The need of approaching the study of structure in its own right, and by independent and (at the start) nonquantitative concepts, is stressed; and a general conceptual model of the structuring of ongoings and events is presented and illustrated in the field of behavior. The combined headings of structural kinematics or geometry (nonquantitative) and structural energics (quantitative) are here found useful. An equation for the latter is developed, and experimental findings thus far obtained in its testing in the field of psychological phenomena are briefly discussed.

REFERENCES

1. Allport, F. H. *Theories of perception and the concept of structure.* New York: Wiley, in press.
2. Hull, C. L., Hovland, C. I., Ross, R. T., Hall, M., Perkins, D. T., & Fitch, F. B. *Mathematico-deductive theory of rote learning.* New Haven: Yale Univer. Press, 1940.
3. McCulloch, W. S. Why the mind is in the head. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior.* New York: Wiley, 1951. Pp. 42–57.

(Received April 9, 1954)

[4] It is hoped that published reports of these studies, presenting the detailed findings and showing the methods of collecting structural information and rendering *E, S,* and *I* operational, will soon be forthcoming. The writer also hopes that further investigations of the hypothesis will be undertaken by others.

Psychological Review
Vol. 61, No. 5, 1954

THE VISUAL PERCEPTION OF OBJECTIVE MOTION AND SUBJECTIVE MOVEMENT[1]

JAMES J. GIBSON
Cornell University

The perception of motion in the visual field, when recognized as a psychological problem instead of something self-evident, is often taken to present the same kind of problem as the perception of color or of form. Movement is thought to be simply one of the characteristics of an object, and the only question is "how do we see it?" Actually, the problem cuts across many of the unanswered questions of psychology, including those concerned with behavior. It involves at least three separable, but closely related problems: How do we see the motion of an object? How do we see the stability of the environment? How do we perceive ourselves as moving in a stable environment?

MOTION, STABILITY, AND MOVEMENT

The first problem concerns the visual perception of a moving object. It seems fairly simple as long as one considers a motionless eye. The stimulus condition for a moving object is the moving sheaf of light rays reflected from it. The retinal image accordingly moves relative to the retina and relative to the background image of the environment. The stimulus for visual movement is retinal movement. This definition is adequate, however, only for a fixated eye. It fails when we consider that the eye normally follows a moving object with a rotary pursuit movement that keeps the image of the object fairly precisely on the fovea. The background image then moves across the retina, but the object image does not. In this case the stimulus for the impression of motion is not so easy to define. A response is going on, and stimulation mediated by this response may enter into the picture. One might assume that movement of the object image *relative to* the background image but not the retina was the effective stimulus. Perhaps the observer senses the motion of the background and perceives the relative motion of the object. Or one might just as well assume that movement of the eye itself relative to the head or relative to the background image but not the object image was the effective stimulus. Perhaps the observer senses the movement of the eye and thereby perceives the motion of the object. The alternatives are highly debatable, but for either one a difficult theoretical question arises: Why do we perceive a motion of the *object* in the environment instead of a motion of the *environment?* This leads to the second problem.

The second problem concerns the visual perception of a stable environment. Why does the world appear motionless, and what are the stimulus conditions for this perception? It is just as much a problem, if less obvious, as the first. Superficially considered, it appears simple for the case of the fixated eye: a motionless image yields a mo-

[1] This research was supported in part by the United States Air Force under Contract No. AF33(038)-22373, monitored by the Perceptual and Motor Skills Research Laboratory, Human Resources Research Center. Permission is granted for reproduction, translation, publication, use, and disposal in whole or in part by or for the United States Government. This report has been based on extended discussions with and contributions from Donald M. Purdy. It also includes ideas and criticisms arising from collaboration with Olin W. Smith.

tionless percept. It becomes difficult, however, for the case of the moving eye. Why does the phenomenal world not move during an eye movement? The eyes perform saccadic or exploratory movements without ceasing during waking life; they perform compensatory movements whenever the head moves; and they perform pursuit movements whenever a moving object catches the attention of the observer. Since the image of the environment moves across the retina during all these responses, the world should seem to move. It may be noted that with certain unusual types of eye movement an observer will report that the world *does* seem to move; examples are the after-nystagmus caused by bodily rotation (or other causes) and the artificial movement of the eye caused by pushing it with one's finger (11). During normal eye movements, however, the world does *not* seem to move, and this poses a question.

There are still other reasons for rejecting the simple hypothesis that a motionless image yields a motionless percept. They appear when we consider what happens when the *observer* moves.

The third problem concerns the visual perception of locomotion in a stable environment. We perceive not only the motions of objects but the movements of ourselves; the performance of fielding a baseball illustrates both. In the case of active locomotion, such as running, there is, of course, a large component of kinesthetic stimulation from the proprioceptors that accompanies the purely visual stimulation from the retinas. But in the case of passive or involuntary locomotion, such as riding in trains, automobiles, and planes, the kinesthetic component may almost wholly drop out. The visual component of stimulation results from the fact of motion parallax, and consists of differential motions of different parts of the image. The writer and collaborators have recently given a mathematical description of this kind of stimulation for the general case of what is called *motion perspective* (8). The fact that it has to do with the perception of space has long been recognized, but the fact that it also has to do with the perception of *locomotion* is less well understood and deserves emphasis. The visual field during forward locomotion seems to expand radially from a point of focus on the line of locomotion. The optical geometry of this expansion is perfectly definite. The retinal image undergoes a deformation that can be neatly specified in terms of differential angular velocities. This retinal motion reaches high magnitudes during rapid travel, and there is reason to believe that it is the important factor in the performance of landing an aircraft. The apparent expansion of the visual field has been noticed by nearly everybody in driving an automobile. The question that arises is why the visual *world* does not seem to expand but instead seems to appear rigid, with the observer moving instead. The flier is never confused by the impression that his runway is behaving like stretched rubber.

It is worth noting that there are special cases of visual stimulation in which it *does* become equivocal whether the visual scene is moving or whether the observer himself is moving. If one sits looking through the window of a stationary railway train at another train on the adjacent track, and if one of the trains begins to move slowly, the impression of moving self with stationary scene may give way to that of stationary self with moving scene, or vice versa.

The three problems of the moving object, the stationary environment, and the moving observer are evidently interrelated. Objective motion is connected with subjective movement,[2] since both stimulate the retina. The motion of an object, the movement of the eye, and the movement of the observer himself may alter the retinal image in different ways, but they all alter it. They are all inseparable from the problem of how or why we see the environment as stationary both when its image is altered and when it is stationary on the retina. One thing is clear at least: the kinetic experience in general involves the problems of so-called space perception.

Experimental Evidence on the Perception of Motion and Movement

A survey of the established facts about the three problems may clarify them and even point to solutions. The experiments are not numerous, some of them are unfamiliar, and they have seldom been considered together.

Motion of an Object

Apparatus employed. Experimental studies of visual perception necessarily depend on devices for systematically presenting light to the eyes of the observer, that is, methods of systematically varying his retinal images. In the case of motion, not many such devices have been successfully built. The types of apparatus for inducing controlled impressions of objective motion are approximately as follows: (*a*) the stroboscope and the variants of this device, used to study apparent motion; (*b*) the moving belt viewed through a window or aperture, used to study so-called "real" motion, or to induce the waterfall illusion; (*c*) the rotating disk with a spiral, used to induce the impression of an expanding or contracting object and the negative afterimage of this impression; (*d*) the device of casting the shadow of a physically moving or rotating object on a translucent screen, the deforming shadow inducing the impression of a three-dimensional object in motion; (*e*) the device of rotating a disk with spiral lines behind a slotted screen, inducing the impression of objects moving along the slot. Practically all that is established about the perception of motion comes from one or another of these experimental methods. However, a novel device for presenting multiple complex motions on a translucent screen has recently been described by Johansson (12). One might suppose that the animated motion picture would have been used for controlled experimentation by psychologists, but it has scarcely been tried (5, ch. 2). There have also been a number of setups with luminous spots in a darkroom, one or more of which are put into relative motion. This latter experiment, like the autokinetic illusion, is relevant to the problem of the stability of the environment as much as it is to the motion of an object.

Stroboscopic motion. The only large body of evidence based on these devices comes from the stroboscope. It is said to yield "apparent" motion as distin-

[2] In this paper, for lack of a better terminology, the word *motion* will always be used to refer to change in position of an object, and the word *movement* will always refer to change in position of the observer's body in whole or part, that is, a response. Both may be visually perceived. The responses with which we are concerned are chiefly eye movements and locomotor movements. Movements of the limbs and hands are also important since they constitute a large part of behavior (gestures, manipulation, tool-using), and most of these are also visually perceived. In them, however, the kinesthetic component, the muscle sense, is obviously important, and the visual component cannot be isolated for analysis as it can for locomotion. They are practically never passive or involuntary, as locomotion can be. They will not be considered here. Nevertheless the writer believes that the visual feedback is just as important for motor performance as the bodily feedback, and that "visual kinesthesis" should be recognized along with classical kinesthesis.

guished from "real" motion, and the stroboscopic effect is often loosely referred to as the phi phenomenon. The stroboscope has evoked much research, probably because it demonstrates that a physically moving object is not necessary for an experience of motion, and because this seeming paradox has prompted psychologists to formulate controversial theories in order to explain it.

The important fact about stroboscopic motion, for present purposes, is that the stimulus is intermittent but that when certain relations hold, the perception of motion is the same as if the stimulus were *not* intermittent. As Troland asserted, "a perfect motion impression can be aroused without any actual motion of an object by the discontinuous substitution of one object for another at progressively different points in space" (18, p. 381). This situation has frequently been reduced for experimental convenience to the case of two successive light sources at two separated points in space, and this experiment has resulted in an elaborate Greek-letter phenomenology of motion impressions (alpha, beta, gamma, delta, and phi). The results of this experiment have been reviewed elsewhere (for example, 1, ch. 15) and will not be discussed here. The fact is that when an adjacent order and a successive order of discrete stimuli are correlated, a continuous impression of an object in motion results. The main limitation seems to be that the interval between stimuli must not be too disproportionate to the separation between them. Hence stroboscopic stimulation differs from so-called "real" stimulation only in being discontinuous when the latter is continuous. The relations of order are the same in both.

Motion of a patterned surface. The speed and direction of linear motion are perceived with some accuracy when a moving belt is presented to the eye. The same thing is true for the rotary motion of the surface of a disk. For both, there are lower thresholds for velocity and also upper thresholds for velocity when motion turns into blur. Acuity for motion is high at the periphery of the retina considering how weak it is for color and form. There occurs a negative afterimage of velocity in a stationary visual field in that part of it which has previously been stimulated by a moving belt or disk. The afterimage may be linear or rotary or it may be one of expansion or contraction if the rotating disk bore a spiral that contracted or expanded (Plateau's spiral). The perceived velocity of a moving surface tends to be constant at different distances from the eye although the retinal velocity of its image is inversely proportional to distance. Brown, however, discovered some other puzzling facts about such apparent velocities connected with the size of the frame or aperture behind which the belt moved and with the brightness of the surface (2). Another fact, which is interesting for the problem of the connection between retinal motion and eye movement, is that perceived velocity is reported to be somewhat faster when the eyes are fixated on the aperture than when they follow the moving pattern from one side to the other and back again. This has been called the Aubert-Fleischl paradox (2).

Deformation of shadows and the perception of depth. Linear and rotary motions presented to the eye by belts or disks occur in the frontal plane of the observer and are so perceived. So does the apparent expansion of a Plateau spiral, and this is also perceived as flat except for an occasional report that the afterimage suggests motion in depth. But the shadow of a rotating object observed from the other side of a translucent screen, although seen in one sense as moving in the frontal plane, is often seen in another sense as moving in depth. There may be a compelling impression of rigid rotation as well as an impression of deformation. This effect has been called *stereokinetic*, and Wallach has recently named it the *kinetic depth effect* (19). Metzger had previously studied the phenomenon and its interpretation (15). The impression of rotation in depth is reversible, and the observation of this feature of it goes back to "Sinsteden's windmill" (1, p. 270).

Controllable complex movements. There have been a few experiments on multiple motions in the visual field, that is, of meaningless spots or shadows moving in systematically varied ways. Michotte, who used the method of a pair of rotated spirals visible through a horizontal slot in a screen, was interested in the perception of causality (16). Metzger, who projected on a translucent screen the shadows of vertical rods rotating on a horizontal turntable, was interested in the problem of the visual identity of the interpenetrating shadows (14). Johansson devised a method of superimposed slide projection in which each spot on the screen depends on a different slide and each slide can be given a controlled linear or circular motion. He was concerned with the perception of the *events* which his moving spots induced (12). Johansson also describes the other important experiments of this type. Heider and Simmel, using animated motion picture film, explored the possibilities of the *social meanings* which moving triangles and circles might evoke (10).

The Stable Environment

In contrast with the foregoing experiments in which the background of the motion, or the frame of the window in which it appears, is always visible stands a class of experiments utilizing points of light in a completely dark room. The case of a single fixated point has been studied for a long time. Although the image remains essentially motionless on the retina of the observer and the spot may appear at first to be static, it eventually shows an "autokinetic" motion. It appears to wander in an erratic fashion, and the observer himself may become disoriented. The illusion disappears if the surfaces of the room become even slightly visible. The facts are summarized by Carr (3, pp. 314 ff.). Evidently the stimulation of a single retinal point is not sufficient to yield the impression of a stable environment. Sandstrom has recently emphasized that an observer cannot even point with his finger to a single spot of light in a dark room (17). Facts of this sort throw great doubt on any kind of theory of the "local signs" of retinal points.

When *two* points of light are presented in the dark, their separation is sensed and they appear connected. They may appear to wander as an autokinetic unit, but one never appears to move relative to the other. It might be said that each has stability relative to the other.

If one of the two point sources in the darkroom is made to move slowly, the conditions are present for what Duncker has called "induced movement" (4). The observer reports motion, but it is as likely to be carried by the physically motionless source as by the moving source. A frequent outcome is a phenomenal motion of both spots, each carrying half of the total velocity. The relative motion of the first to the second or the second to the first (or each to the other) is perceptible, but the motion with reference to the room is not. The room, after all, is invisible and the background of the spots is darkness.

An example of induced motion taken from common experience is the appearance of the moon seen through drifting clouds. In this case the clouds provide an extended background for the moon, not just another spot of light, and the impression of the moon's motion is unequivocal. Duncker set up a similar situation and studied the apparent motion of a stationary spot of light projected on a rectangular surface that moved in pendular fashion from side to side. The relative motion of the spot within the frame was indistinguishable from "real" motion; it could be cancelled by setting up an opposite pendular motion of the spot itself (4).

Duncker also noted the occurrence of induced movement of the observer *himself,* both in the darkroom situation and, under special conditions, with illumination. This was, of course, a movement without kinesthesis, produced wholly by visual stimulation. Insofar as an observer perceives himself in visual space, his own movement, like that of visual objects, depends on the phenomenal frame of reference. The question is, what establishes this frame of reference or stable visual environment?

Movement of the Observer, Including Locomotion

A simple method of inducing by visual stimulation one kind of apparent movement of the observer's body has long been known. It consists of surrounding the head of a stationary observer with a cylindrical screen or curtain, filling his entire visual field, which can then be rotated around the head. The observer reports a perception of being rotated in the opposite direction—an instance of Duncker's "induced ego-movement." The impression may be as vivid as that obtained from being actually rotated in a Barany chair, and the only difference between the case of rotating the miniature visual room and the case of rotating the observer may be the absence of vestibular stimulation in the former and its presence in the latter. The phenomenon is similar in principle to the "railroad train" illusion described earlier.

The analysis of motion perspective for a large portion of the visual field, also mentioned earlier (8), suggests that the impression of *forward* movement of the observer can be produced optically without any contribution from the vestibular or the muscle sense. This experiment, however, has not been performed. The closest approximation to it is an informal study based on a motion picture of the landing field ahead of an airplane during a glide (5, p. 230). Observers reported an experience of locomotion along a glide path toward a visible spot on the ground. This perception was clearly, however, an "as if" kind of experience, pictorial rather than natural. The motion picture intercepted only a part of the field of view. It is said that the panoramic motion picture (especially the "Cinerama") induces even more compelling experiences of locomotion, such as a ride in a rollercoaster.

There has been little or no research on the contribution of kinesthetic, tactual, and vestibular sensitivity to the experience of passive locomotion. Their contribution to the sense of passive rotation of the body has been studied, and something is known about their contribution to the maintaining of upright posture. How kinesthesis is connected with the *visually* aroused impression of locomotion is not known. The flier and the automobile driver have muscular kinesthesis for the *controls* of the vehicle but not for the propulsion of the body, as in walking or running.

The experience of *active* locomotion —of voluntary or guided movement by the observer—is of course a still more complex psychological problem, which will not be touched on in this report. Most of the experimental evidence about voluntary action comes from studies of pursuit tasks, reaction time, and the like, which might be said to deal with manipulation rather than locomotion. A theory of movement with respect to a goal or destination is obviously of great importance, but we are here concerned with the cues or stimuli for movement as such. This may be justified on the grounds that the flow of actions, choices, or decisions during, for instance, an aircraft landing cannot be understood unless the flow of information is understood.

IMPLICATIONS OF THE EVIDENCE

There is plenty of evidence to indicate that visual motion is a "sensory" variable of experience. It has a kind of intensity (speed) and a kind of quality (direction). It has absolute thresholds, both lower and upper, like pitch. Acuity depends on the part of the retina stimulated, like form. It has a negative afterimage, like hue. It tends to manifest constancy, like size and shape. In the form of "pure phi" it can be abstracted from an object. But more than any sensory impression, it *fails to correspond to the physical stimulus presumed for it.* Whatever the stimulus for motion might be, it is *not* simply motion in the retinal image. This seems to imply that motion is not sensory. Before concluding, however, that phenomenal motion is not a function of stimulation, the stimulus conditions should be re-examined.

The distinction between "real" and "apparent" motion is unfortunate and has interfered with the search for the essential conditions. It should be noted that stroboscopic stimulation can yield just as psychologically "real" a motion as does continuous stimulation, if certain relations are preserved. A stroboscope and a moving object are manifestly different, but they are the *sources* of stimulation, not the stimuli, and perhaps the latter are not so different after all. The facts of the experiments can be explained by the hypothesis that the retina responds to adjacent and successive *order.* If the orders correlate for the stroboscope and the object, the fact that the former is a discontinuous emitter may be unimportant. The two retinal images are similar in that the relations of order are the same in both; for example, right-left and before-after. The stimulus for motion, then, may be *ordinal.*

There is other evidence to suggest that the stimulus for motion is also *relational.* This means that it cannot be derived from the hypothetical "local signs" of retinal receptors. The fovea does not have a fixed value for breadth and height when stimulated by a single point of light. Moreover, as Duncker proved, the motion of one point of light on the retina is perceived relative to another point of light, not relative to the retina. The frame of reference for motion (or stability) seems to depend on the array of stimulation rather than the location of the receptors; it is transposable over the retina. Just as a motion for the physicist can be specified only in relation to a chosen coordinate system, so is a phenomenal motion relative to a phenomenal framework (13). Perceived motion occurs in a perceptually stable space or environment. Another way of saying this is to assert that the perception of stability is part and parcel of the perception of motion; you cannot have the latter without the former.

The optical stimulus conditions for a stable environment seem to be a retinal image containing many elements rather than a few or one. This can be described as a differentiated or "textured" image (7, 9). Perhaps stability goes with the perception of a surface or an array of surfaces extending over most of the field of view. The disappearance of the autokinetic illusion when the darkroom is even slightly illuminated is consistent with this hypothesis. So is the occurrence of the moon-in-the-clouds illusion. So also is the railroad train illusion when we take the window-filling train on the next track to be motionless. Perhaps *the textural background image, whatever its relation to the anatomical retina, always tends to determine the phenomenal environment, and the more it approximates the total image the greater the stability.*[3]

Common experience suggests that we can perceive the motion of an object in depth as readily as its motion at right angles to the line of sight, and the experiments with deforming shadows on a translucent screen tend to bear out this suspicion. The kinetic depth effects so far obtained depend on perspective transformations of the shadows, and yield impressions of changing slant or rotation. There is no reason why they should not also be obtained with size transformations of shadows, which will yield impressions of linear approach and recession. A general hypothesis is suggested by these experiments, namely, that *any regular transformation of a bidimensional image tends to yield a tridimensional motion in perception, and the kind of motion perceived depends on the kind of transformation.* This hypothesis has the advantage of relating the experiments on moving shadows to experiments on shape constancy and size constancy, and suggests a principle of space perception that may be common to both. The fact that the transverse motions of a pair of belts observed at different distances can be judged equal in velocity when the surfaces are actually equal in velocity (if Brown's results [2] are accepted) points in the same direction.

Facts about the perception of bodily movement as distinguished from object motion are scarce. They are enough to suggest, however, that the impression of oneself being moved, like that of an object being moved, depends on the perception of the space in which the movement occurs. Ego movement like object movement can be induced. The train illusion and the cylinder rotating around the head are examples. The perception of forward locomotion can probably be induced, and the experiment should be tried. This will require optical stimulation governed by differential angular velocities for many points in the visual field, i.e., motion perspective or, crudely speaking, an expanding image.

A promising hypothesis for research would be that *any transformation of the total retinal image, as distinguished from a part image within it, tends to yield an experience of a movement of the observer, and the kind of movement experienced depends on the kind of transformation.* For example, a simple translation of the image may contribute to the experience of an eye movement; an expansion may contribute to the experience of forward locomotion; a contraction to the experience of backward locomotion; and so forth.

There is said to be a striking lack of correspondence between the presumable optical stimuli and the ensuing visual perceptions of motion or movement. The evidence does indeed show what appear to be obvious discrepancies. It is certainly true that kinetic impressions are not *copies* of their stimuli. But it fails to follow that they are not *functions* of their stimuli. It cannot simply be assumed that a movement is the same thing in the object, the retina, the brain, and consciousness. The foregoing hypotheses make it possible to

[3] This hypothesis is consistent with, if not essentially the same as, the position taken by Duncker in his admirable study of "induced" movement (4).

test for psychophysical correlations, although they do not imply any pictorial correspondence, between the dimensions of the stimulus and the qualities of kinetic experience.

Hypotheses about Kinetic Retinal Stimulation

A psychophysics of kinetic impressions would require a mathematical analysis and classification of the motions or transformations of a retinal image. This is a complex and difficult task for the future. Some preliminary assumptions are possible, however.

Geometrically, one can distinguish between a *rigid* and a *nonrigid* motion of a form or of a set of points. *Translation* and *rotation* are the types of rigid motion with which we are concerned. The figure after displacement is congruent or identical with the figure before displacement. The kinds of nonrigid motion are diverse and are still being explored by the higher branches of geometry. However, two classes exist, which may be called *elastic* motion and discontinuous or *disjunctive* motion. In the former, the lines of the geometrical form do not "break up" (or the set of points maintains the relations of neighborhood), whereas in the latter the form is ruptured(or the points are "scattered"). The class of elastic motions includes two types, the *size transformations* and *perspective transformations* on the one hand and *nonperspective transformations* on the other. The first type can be defined as a projection of the form or pattern on a plane different from its own, either an enlargement (or reduction) or a slant projection. The second type can be defined as a deformation other than these, but for which the continuity of the form is preserved. The class of disjunctive motions includes many types, which do not need to be specified here, but all involve discontinuity. The six types with which we are concerned are tabulated below:

Rigid motion
1. Translation
2. Rotation

Elastic motion
3. Size transformation
4. Perspective transformation
5. Deformation

Disjunctive motion
6. Multiple movements

These abstract mathematical motions are interestingly related to optical stimulation. Let us assume an eye and a reflecting surface, such as the face of an object toward the eye, and let us consider the cross-section of the sheaf of light rays to the nodal point of the eye (18, pp. 326 f.). This is equivalent to the retinal image. What tridimensional events produce these motions of the bidimensional cross-section? Numbers 1 and 2 above correspond respectively to a lateral movement of the eye (or the object) and a swivel movement of the eye (or a rotation of the object). Number 3 corresponds to a movement of the eye (or object) along the line between them. Number 4 corresponds to a planetary movement of the eye around the object or an inclination of the object to the line between them. Number 5 corresponds to an event confined to the object—a fluid or elastic motion of its substance. Finally, number 6 probably corresponds to an event such as the shattering of a single object or the interaction of multiple objects. Some of these statements need qualification in order to be exact, but they may serve as preliminary general rules. In other words, some very important types of physical events correspond to the geometrical types of motion in the projection. It is a reasonable hypothesis that the eye can *register* these geometrical types of motion when they occur in the retinal image.

It may have been noted that the physical events corresponding to motions number 1, 2, 3, and perhaps 4 are ambiguous. Whether the eye moves or the object moves, the result is the same. The optical situation assumed in the previous paragraph consisted of an eye and a single object (specifically a plane *face* of an object). A more typical optical situation would consist of an eye and an *environment*. Let us therefore assume instead an eye and an infinite plane surface. This is a better approximation to the terrestrial environment. Except for the "sky," the image of the surface occupies the whole of the retina, and it constitutes a textured background image rather than a delimited object image. An infinite plane surface would be physically stable and would constitute an excellent frame of reference for visual perception. There is evidence to suggest that a background image does help to determine the stable phenomenal environment. Ambiguity of perception as to whether the eye moves or the environment moves in *this* situation would therefore tend to disappear.[4]

The types of physical events producing the geometrical types of motion of a total background image are fairly univocal. Translation and rotation of this image can hardly be caused by anything but eye movements. Size and perspective transformations for the elements of an extended plane surface constitute motion perspective (8) and this can hardly be caused by anything but locomotion with respect to the surface. Certainly it is true that any eye movement in an illuminated environment causes a rigid movement of the image, and any transportation of the eye causes an elastic movement of the image.[5]

The causes in the environment and the results in perception of deformations and disjunctive motions of the image (numbers 5 and 6 above) are complicated. So far, we have been assuming a solid environment. Nonperspective deformations are caused by liquid or fluid motions of physical objects and surfaces. Rivers flow, smoke swirls, rubber stretches, and above all living organisms flex their surfaces in many ways. The faces of men, for instance, undergo an astonishing variety of rubbery motions, which we call facial expressions. We perceive these motions, sometimes with great acuity. We do not seem to confuse them with the mechanical motions of solid objects which tilt, slant, advance, or recede with a kind of inanimate quality. There may be a basis in optical stimulation for this difference.

Disjunctive motions of the image are caused by a still greater variety of events. Objects break, ants swarm, billiard balls collide, and men shake hands. Michotte believes that multiple motions can yield immediate impressions of causation that are specific to the relations between them, and he has fortified his belief by experiments (16). The possibility of isolating high-order variables of stimulation in such images seems remote, but it should not be rejected.

In conclusion, the various motions of objects in a stable environment and the various movements of ourselves in that environment can both be visually perceived. A psychophysics of such kinetic impressions, however, is almost nonex-

[4] If to our disembodied eye we add assumptions about gravity, posture, muscles, and kinesthetic stimulation, the ambiguity would certainly disappear. But we are here concerned only with optical stimulation, admittedly an abstraction.

[5] The classification of the motions of a retinal image here given is considerably revised from that proposed previously by the writer (6, p. 131 ff.).

istent, and the possibility of isolating their stimuli has been doubted. If, however, the effective stimulation is taken to be ordinal and relational, it falls into several mathematical classes, which are neatly correlated with types of physical events, and which may prove to be psychophysically correlated with modes of kinetic experience.

REFERENCES

1. BORING, E. G. *Sensation and perception in the history of experimental psychology.* New York: D. Appleton-Century, 1942.
2. BROWN, J. F. The visual perception of velocity. *Psychol. Forsch.*, 1931, 14, 199–232.
3. CARR, H. *An introduction to space perception.* New York: Longmans, Green, 1935.
4. DUNCKER, K. Über induzierte Bewegung. *Psychol. Forsch.*, 1929, 12, 180–259.
5. GIBSON, J. J. (Ed.) Motion picture testing and research. Washington, D. C.: Government Printing Office, 1947. (*AAF Aviat. Psychol. Program Res. Rep.* No. 7.)
6. GIBSON, J. J. *The perception of the visual world.* Boston: Houghton Mifflin, 1951.
7. GIBSON, J. J., & DIBBLE, F. N. Exploratory experiments on the stimulus conditions for the perception of a visual surface. *J. exp. Psychol.*, 1952, 43, 414–419.
8. GIBSON, J. J., OLUM, P., & ROSENBLATT, F. Motion parallax and motion perspective in aircraft landings. *AF Hum. Resour. Res. Cent. Bull.*, in press.
9. GIBSON, J. J., & WADDELL, D. Homogeneous retinal stimulation and visual perception. *Amer. J. Psychol.*, 1952, 65, 263–270.
10. HEIDER, F., & SIMMEL, M. An experimental study of apparent behavior. *Amer. J. Psychol.*, 1944, 57, 243–259.
11. HOLT, E. B. Eye-movement and central anaesthesia. *Psychol. Monogr.*, 1903, 4, No. 1 (Whole No. 17), 3–46.
12. JOHANSSON, G. *Configurations in event perception.* Uppsala: Almqvist & Wiksell, 1950.
13. KOFFKA, K. *Principles of Gestalt psychology.* New York: Harcourt, Brace, 1935.
14. METZGER, W. Beobachtungen über phänomenale Identität. *Psychol. Forsch.*, 1934, 19, 1–60.
15. METZGER, W. Tiefenerscheinungen in optischen Bewegungsfeldern. *Psychol. Forsch.*, 1935, 20, 195–260.
16. MICHOTTE, A. *La perception de la causalité.* Louvain: Inst. sup. de Philosophie, 1946.
17. SANDSTRÖM, C. I. *Orientation in the present space.* Uppsala: Almqvist & Wiksell, 1951.
18. TROLAND, L. T. *Principles of psychophysiology.* Vol. 1. *Problems of psychology and perception.* New York: Van Nostrand, 1929.
19. WALLACH, H., & O'CONNELL, D. N. The kinetic depth effect. *J. exp. Psychol.*, 1953, 45, 205–217.

(Received November 16, 1953)

Psychological Review
Vol. 61, No. 5, 1954

VARIABLES AND FUNCTIONS[1]

ABRAHAM S. LUCHINS AND EDITH H. LUCHINS

University of Oregon

Psychologists use such mathematical terms as variable, independent and dependent variable, and function, as well as the mathematical symbolism for representing functional relationships. But they often fail to specify whether or not these terms and symbols have the same meanings as in mathematics, and thereby, it seems to us, pave the path for confusion. There follows an attempt to highlight some of the differences between mathematicians' and psychologists' uses of these terms and symbols.

What Is A Variable?

Mathematical texts generally define a variable as a symbol that may represent different objects during the course of a given discussion. The set of objects with which the variable may be identified is called the variable's range or domain. For example, if T represents the set of all integers, and if the domain of variation of X is T, then X may be considered to represent any arbitrary integer. The number of objects in the domain of a variable may be limited (e.g., the set of all numbers from 1 to 10) or unlimited (e.g., all even integers). Mathematicians have found it convenient to include a constant, a symbol restricted to represent one object during the course of a particular discussion, as a special case of a variable by thinking of a constant as a variable whose domain of variation consists of only one element.

Psychological texts do not usually state explicitly what they mean by the term *variable*. One recent report, which does offer a definition, effectively eliminates the possibility of considering a constant as a special case of a variable. The report requires of a variable that it *vary*, that it assume at least two values, and accordingly defines a variable as "a set of two or more categories such that, if any object or event be a member of one of those categories, it may not be a member of any other of those categories" (1, p. 46). Psychologists generally have not indicated whether they would consider a constant as a case of a variable. But some have done so implicitly; for example, in speaking of "the stimulus variable," they intend reference to only *one* object, say, a lever.

There is some difference of opinion among psychologists as to the major *kinds* of variables. Some psychologists recognize two kinds of variables, stimulus variables and response variables, or, as Spence (11) describes them, S variables and R variables. Other psychologists speak of three kinds of psychological variables: stimulus, response, and organismic variables (7); or variables pertaining to behavior, to environment, and to individual differences (13). In addition, some, but not all, psychologists regard intervening constructs (or logical constructs) as another kind of variable. Still others divide variables according to whether or not they are experimentally manipulable.

It is not clear whether this division is intended to constitute a stipulation of the range of variation of a given variable. It is a serious shortcoming of psychologists' usage of the variable concept that the range of variation is usually not clearly specified. While the mathematician stipulates that the vari-

[1] This is a condensation of a chapter to appear in the authors' book dealing with a variational approach to psychology.

able *X* is limited to integers or to real numbers, for example, or that it may range over the entire complex number domain, and while he may explicitly eliminate certain objects in the range (e.g., all values of the variable that would introduce division by zero [10, p. 99]), the psychologist does not usually indicate what is to be included in, or excluded from, the domain of variation. Accordingly, the psychologist may speak of the stimulus variable without specifying whether its range includes all stimuli or only one stimulus or a certain restricted class of stimuli. Similarly, he may refer to the response variable as if it were permitted to range over the set of all possible responses or at least all responses made by the subject during the experimental session, and yet, in actual practice, he may identify the variable only with the time required to trace a maze or with the frequency of lever pressing within a given time, thereby effectively ruling out many "responses," e.g., those referring to the pressure on the lever or the member of the body which pressed the lever. Perhaps it would help to decrease confusion if psychological writings made a point of explicitly mentioning the domain of the variable.

There is also a difference of opinion as to whether quantification is prerequisite for a variable. Some psychologists hold to the belief that a variable should properly be "quantifiable," that it should pertain to things measurable: to numbers. Only if such quantification is a characteristic do they speak of a variable. Bergmann and Spence (2) write that when a relevant factor underlying a phenomenon *has become quantifiable,* it is called a variable. But other psychologists (e.g., 7, p. 4) claim that many of the variables used in psychology are qualitative and not quantitative.

Even those who permit qualitative variables in psychology often express the opinion that they are less preferable than quantitative variables and that the aim should be to introduce measurements whenever possible. In short, the most desirable domain of variation in psychology would seem to be a domain of numbers. If we use the phrase "numerical variable" to refer to a variable whose domain of variation consists of numbers, then we may sum up this state of affairs by noting that psychologists currently tend to favor, or even to advocate a restriction to, numerical variables.

This is not the case in mathematics. By no means are all the variables with which the mathematician deals numerical variables (5, p. 273). He is concerned also with nonnumerical variables, with those whose domains of variation do not consist of numbers. He may be dealing with a variable that ranges over the set of all triangles in the plane, or the set of all curves in the plane joining two given points, or the set of all minor arcs on a sphere, or the set of all properties remaining invariant under a given transformation, etc. It may be retorted that some of these domains of variation consist of objects that are "quantifiable." For example, one might measure the area of a triangle or the length of a curve. But this is quite beside the point. When a variable has as its range, say, the set of triangles in the plane, then the variable is identified with a *triangle,* and not simply with any or every "quantifiable" or "numerical" aspect of it such as the area of the triangle, or its perimeter, or the heights of its altitudes, or the numerical values of its interior angles, or the lengths of its angle bisectors or medians. The range of variation is the set of triangles in the plane, and not the set of quantifiable or measurable aspects of the triangles.

Possibly the current trend in psychology toward quantification and the emphasis on numerical variables stem from the prevalent belief that science typically strives to quantify its constructs (11). Whether or not this is the case, it may be a comfort to psychologists who are uneasy about "qualitative" variables to keep in mind that mathematics, as a tool of the sciences, is equipped to deal with certain qualitative relationships, and that mathematicians (including mathematical physicists) freely work with nonnumerical variables and neither regard them as inferior to numerical variables nor insist upon transforming them into the the latter.

What Is A Function?

If to each value of a variable X with a range M, there is associated one or more values of a variable Y with a range N (where M may or may not be identical with N), then Y is called a mathematical function of X. This is usually written symbolically as $Y = f(X)$. If only one value of Y belongs to each value of X, Y is called a single-valued function of X and, otherwise, a multiple-valued function. It may happen that to a variable Y there are associated values of a *set of variables*, $X_1, X_2 \ldots X_n, \ldots$, with specified ranges in which case $Y = f(X_1, X_2, \ldots X_n \ldots)$.

Psychologists make frequent use of the terminology and symbolism of mathematical functions. But they sometimes use these in a manner that differs radically from mathematical usage. For example, in psychological discussions we may find the concept of function, represented by such symbolism as $R = f(S)$, where R represents response and S represents stimulus, interpreted implicitly or explicitly to mean a causal relationship, e.g., that S is the *cause* and R the *effect* of this cause. But a mathematical function does not (in mathematics at least) denote cause and effect.

In Courant and Robbins (5), we find the following:

> A mathematical function is simply a law governing the interdependence of variable quantities. It does not imply the existence of any relationship of "cause and effect" between them. Although in ordinary language the word "function" is often used with the latter connotation, we shall avoid all such philosophical interpretations. For example, Boyle's law for a gas contained in an enclosure at constant temperature states that the product of the pressure p and the volume v is a constant c (whose value in turn depends on the temperature): pv = c. This relation may be solved for either p or v as a function of the other variable, . . . without implying that a change in volume is the "cause" of a change in pressure any more than that the change in pressure is the "cause" of the change in volume. It is only the form of the *connection* between the two variables which is relevant to the mathematician (5, p. 276).

Psychologists have also employed the function concept to represent a prior-subsequent relationship so that $R = f(S)$ is interpreted as implying that S is prior and R subsequent to S. This has led to controversy as to whether it is the stimulus or the response which enjoys priority (cf. 1, p. 48). But a mathematical function is not a representation of prior-subsequent relationships any more than it represents cause-and-effect relationships. For example, the functional relationship between areas and radii of circles implies neither the priority of the radius nor that of the area.

Mathematicians have found it convenient to conceive of a function, $Y = f(X)$, as a mathematical operation $f(\ )$ which, when applied to X yields Y, or as a mapping or transformation of one domain into another, in this case the domain M of the variable X into the domain N of the variable Y. It has been pointed out that mathematicians usually stress the form of the connection between the variables, the "law of correspondence" or the operation $f(\ )$, while physicists are usually more interested in the *result* of the operation; "confusion can sometimes be avoided only by knowing exactly whether one means the operation $f(\ )$ which assigns to X a quantity $u = f(x)$, or the quantity u itself, which may also be considered to depend, in a quite different manner, on some other variable, z" (5, p. 277). In psychology, too, confusion can sometimes be avoided if in a given context the referent of the term *function* is indicated. For example, in dealing with $R = f(S)$, psychologists may unwittingly confound the discussion if they do not specify whether they are using *function* to designate the operation $f(\ )$ or R itself.

Finally, we should like to refer to the current stress in psychology on explicit formulations of functional relationships, on determination of the exact mathematical formula connecting variables. This is well exemplified in the writings of Hull (8, 9) and Spence (11). Spence writes, "Instead of knowing merely that the response, R, is some function of the variables $X_1, X_2, X_3 \ldots X_n$, he [the psychologist] desires to know the precise function." Hull is highly critical of unstated (nonexplicit) functional relationships and argues for explicit determination of the mathematical formula connecting the variables (8, p. 29). He contends that it is better to make an incorrect guess as to the exact nature of the relationship than to work with nonexplicit relationships, and calls for "determination of what the function actually is" (9, p. 173).

This emphasis may be related to the belief that in order to characterize a mathematical function it is necessary to have an explicit statement of the exact mathematical formula connecting variables. But this involves a notion of function and of functional relationship that has proven *too narrow* for the needs of higher mathematics.

> To Leibniz (1646–1716), who first used the word "function," and to the mathematicians of the eighteenth century, the idea of a functional relationship was more or less identified with the existence of a simple mathematical formula expressing the exact nature of the relationship. This concept proved too narrow for the requirement of mathematical physics, and the idea of a function . . . was subjected to a long process of generalization and clarification (5, p. 273).

Note that the first impetus for a broader conception of the idea of a function came from mathematical physics, that example par excellence of the fusion of problems of a science with mathematical methods. If behavior scientists are to be successful in evolving a "mathematical psychology," it may be necessary that they also discard a too narrow view of the idea of a functional relationship. In any event, the generalization of the idea of function is by no means confined to mathematical physics or to physics but is common to all of higher mathematics. Often the mathematician is concerned with functional relationships for which he does not have a simple, or even not-so-simple, mathematical formula that precisely stipulates the nature of the relationship. But this need not be a handicap to further mathematical treatment since the modern conception of function does not require that it be expressed in terms of a combination of elementary functions such as logarithms, exponentials, sines, etc., or that it be expressed as a power series, as a trigonometrical series, or in terms of derivatives or integrals. So long as to each value of the independent variable or to the set of independent variables there corresponds a value of the dependent variable, then the latter is said to be a function of the former despite the lack of any more precise mathematical formulation. The mathematician may therefore prove (or assume) that there exists a functional relationship, which he may express in the general form, $u = u(X_1, X_2, \ldots X_n)$, that the variables have certain ranges, and that certain properties are possessed by the function; for example, that it is differentiable or has derivatives up to a certain order, or that it is continuous or semicontinuous. And yet from such knowledge (or assumptions) he may be able to derive many further mathematical properties and relations. Not being able to express the exact nature of the relationship by a simple mathematical formula is therefore not a barrier to mathematical treatment, either in mathematics itself, or in physics, or in other applications of mathematics.

Likewise, ignorance of the exact nature of the mathematical formula connecting psychological variables need not prove an insuperable obstacle to further mathematical treatment. Hull suggests that the psychologist should be prepared to think in terms of higher mathematics (8, p. 400). But then he must be prepared to accept the broader notion of function common to higher mathematics, and not to be uneasy about "unstated" functional relationships if by this he means that the explicit formula is not revealed. To think mainly in terms of explicit formulas, to regard their determination as the proper aim of the psychologist (even to the extent that an incorrect guess as to the specific nature of the relationship is regarded as scientifically sounder than working with a nonexplicit relationship), may be the consequences of a narrow conception of function. To insist on this conception is to bind psychology in the swaddling clothes kicked off by mathematical physics and other branches of higher mathematics during their infancy.

INDEPENDENT AND DEPENDENT VARIABLES

If $Y = f(X)$, it is conventional to speak of Y as the dependent variable and X as the independent variable since the value of Y is dependent on the particular value of X chosen. The number of independent variables may be one, as in $Y = f(X)$; or two, as in $Y = f(X, Z)$; or n, as in $Y = f(X_1, X_2, \ldots X_n)$; or infinite, as in $Y = f(X_1, X_2, \ldots)$.

It is worthwhile emphasizing that the designation of one variable as the dependent variable, and another as the independent variable, may be nothing more than a convention. There may be nothing inherent in the nature of the variables that makes one dependent on the other and not vice versa. For example, in the case of a single-valued function of one variable, $Y = f(X)$, there always exists the unique *inverse function*, $X = g(Y)$. If $f(\ )$ is the function which maps a domain M into a domain N, then the inverse function $g(\ )$ is the one which maps N into M. For the $X = g(Y)$, convention decrees that X be called the dependent variable and Y the independent one, thus reversing the labels attached to these variables when $Y = f(X)$. Hence whether Y or X is labeled as the independent variable may depend solely on the manner in which the functional relationship is written.

The nomenclature, dependent and independent variable, in mathematics is therefore not construed to mean that one variable is completely independent of the other. The very fact that the nomenclature presupposes a functional relation means that the so-called dependent and independent variables are actually interdependent. The function concept, we have seen, is a "law governing the interdependence of variable quantities," and this interdependence of course works both ways. Any change in the dependent variable involves a change in the independent variable, and the other way around, precisely because the two are *connected* by a functional relation. For example, a change in the area of a circle is concomitant with a change in the radius, and vice versa.

Thus the word *independent* in this context has a meaning that is different from that associated with the word in everyday parlance. It is also different from the meaning associated with the term when one refers to a system with n degrees of freedom as having n independent variables or n independent coordinates. These n variables (or coordinates or degrees of freedom) are actually independent of one another. Attempts to describe the system by means of fewer, say $n-1$, variables, would prove inadequate; and if $n+1$ variables were introduced, one variable would prove to be a linear combination of the others (to be linearly dependent on the others). If a system has n independent variables (n degrees of freedom), a change in one variable does not presuppose a change in the other variables. Indeed, it is precisely the ability of the n coordinates to vary independently of one another which leads to their being described as degrees of freedom and as independent coordinates. But, we reiterate, it is not in this sense that the term *independent* is to be interpreted when discussing the independent and dependent variables of a functional relationship.

Perhaps the nomenclature of *dependent* and *independent* in the latter context represents an unfortunate choice, particularly since it seems to have produced some misinterpretations among psychologists. At least, the misinterpretation seems to be sufficiently widespread to lead Bakan to state in a recent report:

> What is generally meant by the assertion of the independence of the stimulus is that it is independent of the response. The paradigm is that the response is dependent on the stimulus, but the stimulus is independent of the response. The formulation of this paradigm is R = f(S) (1, p. 48).

Bakan himself is opposed to this interpretation. Whether the interpretation is as common as Bakan claims or is limited to only a few psychologists, there would seem to be a need to reevaluate what is meant by the functional notation relating S and R, and by their "independence" and "dependence," respectively.

Presumably $R = f(S)$ is intended to convey the idea that R is a mathematical function of S. At least, the notation is that of the mathematical function, and most psychological texts imply that such a function is involved. But if the stimulus and response are related by a mathematical function—by any law, rule, or correspondence that associates with each value of the stimulus one or more values of the response—it follows that the stimulus and response are interdependent and that one cannot properly speak of one as independent of the other. If the paradigm is actually intended to be that the stimulus is independent of the response but the response dependent on the stimulus, then it would perhaps be best not to use the notation of a mathematical function.

We are not objecting to the description of the stimulus as the independent variable and of the response as the dependent variable since it is conventional to attach the labels in this manner when the notation $R = f(S)$ is used. But we are taking issue with the interpretations of the descriptive adjectives.

Bakan considers it inappropriate to talk of the independent stimulus variable because "the stimulus does not exist as a stimulus except by virtue of the responses of the organism" (1, p. 48).

Interestingly enough, the same issue of the journal containing Bakan's report also includes another article, which calls for concepts immediately physical or reducible to physical concepts, and which maintains that under this canon a stimulus "would have to be something which an experimenter could ascertain without there being any organism for it to work on" (6, p. 10); that is, a stimulus would exist when there was no organism to respond to it. While Bakan does not deal specifically with this contradiction to his own point of view, he holds that it is because of a mistaken notion of the priority of the stimulus that some psychologists speak erroneously of the independence of the stimulus and the dependence of the response variable. On the other hand, he continues, some of the very persons who talk of the independent stimulus variable (e.g., 2, 4, 12) give priority not to the stimulus, but to the response. Thus he notes that Stevens (12) accepts the priority of the operation but goes on further to specify the nature of the operation as being *discrimination,* a response.

Another possible source of the interpretation that some psychologists give to independent and dependent variables may be the belief that the former is the cause and the latter the effect, that the stimulus, for example, is the cause and the response its effect or, more generally, that the independent variables are the "initiating causes of behavior" and the dependent variables the resulting behavior.

Psychologists are, of course, free to assign priority to the stimulus or to the response or to assume a causal relationship connecting them. But then the use of the mathematical function notation may be misleading since, as we have already indicated, this notation (at least in mathematics) does not imply a prior-subsequent or cause-effect relationship.

SUMMARY

While psychologists use such mathematical terminology as *variable, independent* and *dependent variable,* and *function,* as well as the mathematical symbolism for representing functional relationships, they often fail to specify whether or not the terms and symbolism are to be interpreted as in mathematics. Other interpretations are sometimes implicit or explicit in psychological writings. The present report outlines some of the differences between mathematicians' and psychologists' use of these terms and symbols.

1. The term *variable* is generally left undefined in psychological texts. Whether or not a constant is accepted as a special case of a variable (as it may be in mathematics) is usually not indicated.

2. Psychologists often speak of a variable (e.g., stimulus variable or response variable) without indicating what is to be included in, or excluded from, its range of variation, that is, without indicating the set of objects with which the variable may be identified.

3. Some psychologists consider quantification or measurement essential to (or at least preferable for) a variable, suggesting that the only (or the preferred) range of variation should consist of numbers. But mathematicians work with both numerical and non-numerical variables (those whose domain of variation does not consist of numbers) and do not consider the latter inferior to the former or as requiring reduction to numerical variables.

4. The mathematical function notation is sometimes employed by psychologists to imply that one variable is prior to another in time or that one variable is the cause and another the effect. In mathematics the term *func-*

tion does not imply either a prior-subsequent or a cause-effect relationship.

5. Confusion may be avoided if psychologists indicate whether, in a specific context, they mean by the term *function* the mathematical operation $f(\ )$ (which may also be interpreted as a mapping or transformation) or the outcome of applying $f(\ )$ to a variable or to a set of variables, that is, whether, for $Y = f(X)$, the referent of the term *function* is $f(\ )$ or Y.

6. We pointed out that the so-called independent and dependent variables of a functional relationship are actually mutually interdependent (since they are related by a function concept), so that it is erroneous to assume, as some psychologists apparently have, that one of the variables is independent of the other, but not the reverse. Whether a variable is designated as dependent or independent may be a consequence of the manner in which the functional relationship happens to be written, and may not be an indication of anything inherent in the nature of the variable.

7. Different interpretations of the relation $R = f(S)$, where S represents stimulus and R response, were analyzed, and the basis of current controversies was found to lie in the conflicting interpretations given to the terms *function, independent variable,* and *dependent variable.*

8. It was suggested that the current stress in psychology on explicit formulation of functional relationships, on determination of the exact formula or equation, may involve a notion of function that has proven too narrow for the needs of higher mathematics, and that may be too narrow for psychology.

REFERENCES

1. Bakan, D. Learning and the scientific enterprise. *Psychol. Rev.*, 1953, 60, 45–49.
2. Bergmann, G., & Spence, K. W. Operationism and theory in psychology. *Psychol. Rev.*, 1941, 48, 1–14.
3. Bergmann, G., & Spence, K. W. The logic of psychophysical measurement. *Psychol. Rev.*, 1944, 51, 1–24.
4. Bridgman, P. W. *The logic of modern physics.* New York: Macmillan, 1938.
5. Courant, R., & Robbins, H. *What is mathematics?* New York: Oxford Univer. Press, 1941.
6. Davis, R. C. Physical psychology. *Psychol. Rev.*, 1953, 60, 7–14.
7. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
8. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
9. Hull, C. L. Behavior postulates and corollaries—1949. *Psychol. Rev.*, 1950, 57, 173–180.
10. Richardson, M. *Fundamentals of mathematics.* New York: Macmillan, 1941.
11. Spence, K. W. The nature of theory construction in contemporary psychology. *Psychol. Rev.*, 1944, 51, 47–68.
12. Stevens, S. S. Psychology and the science of science. *Psychol. Bull.*, 1939, 36, 221–263.
13. Tolman, E. C. Operational behaviorism and current trends in psychology. *Proc. 25th Anniv. Celebr. Inaug. Grad. Stud.* Los Angeles: University Southern California Press, 1936. Reprinted in M. M. Marx (Ed.), *Psychological Theory*, New York: Macmillan, 1951. Pp. 87–102.

(Received December 4, 1953)

Psychological Review
Vol. 61, No. 5, 1954

BEHAVIOR UNDER STRESS: A NEUROPHYSIOLOGICAL HYPOTHESIS[1]

H. RUDOLPH SCHAFFER[2]
Tavistock Clinic, London

Stress and its effects on behavior is a subject of which both clinicians and experimentalists have become increasingly aware within the last two or three decades. In human beings it has been observed mainly under naturally occurring conditions, as instanced by the literature on psychiatric war casualties (e.g., 18) and, in the case of infants, by the experience of prolonged separation from the mother, which Bowlby (4) has found to have marked pathological effects on development. In animals stress has been studied chiefly in the laboratory, where a variety of techniques has been evolved to disrupt existing adaptation patterns and to replace these by certain forms of nonadjustive behavior. Probably the fullest and most systematic set of data derives from those studies which have been grouped together under the title of "experimental neurosis," and it is to this work that the present paper will primarily refer.

Our knowledge of the behavioral data and of many of the antecedent conditions important in this type of stress situation is now, thanks to a considerable number of descriptive studies, fairly adequate, but no completely successful attempt has yet been made to fit a theoretical framework to the results of this work. It is the aim of this paper to propose such a framework, and it is hoped that, with the help of certain neurophysiological notions, it will thereby become possible to order the behavioral data and to reach some understanding of the dynamics underlying behavior under stress.

Stress as Disruption of the Organism-Environment Relationship

A stressful situation may be regarded as essentially one in which a major disruption of the relationship between an organism and its environment has taken place. Under nonstressful conditions this relationship tends to be relatively harmonious; the environment, on the one hand, will gratify the needs and expectations of the individual, while the organism, on the other, can adequately meet the demands made upon it by external stimulation. The relationship will suffer some disruption when the organism meets a novel situation for which it has no adjustive response readily available and in which it cannot find such a response until a period of trial-and-error behavior (problem solving) has taken place; but even in such circumstances, the disruption will generally be a minor one and adjustment will eventually occur. It is only when the difficulty of the task confronting the organism is so increased as to approach a no-solution situation that one may speak of a stress situation.

But the relation between the complexity of a problem and the capacities of the individual to solve it is only one factor defining the organism-environ-

[1] This paper was written as part of a project of research on child development undertaken at the Tavistock Clinic, London, and sponsored jointly by the International Children's Center, Paris, and the European Division of the World Health Organization.

[2] The author wishes to express his sincere gratitude to Dr. J. Bowlby for his encouragement and advice, and to Professor J. Elkes and Dr. M. D. Ainsworth for many helpful suggestions.

ment relationship. Another is the degree of motivation which impels the organism to face the problem, and which thus restrains it within the situation. Only when a drive has been activated can we talk of a psychological relationship between the individual and his environment, and only when this drive is sufficiently strong to persist is it possible for a major disruption in this relationship to occur. Thus, as Lazarus, Deese, and Osler (20) have pointed out, stress cannot be defined in terms of the environment alone; the motives and capacities of the organism and their interaction with the environment must also be taken into account. Fuller's (11) definition of a stressful situation as one in which adjustment is difficult or impossible but in which motivation is very strong satisfactorily complies with this demand.

Disruption of the organism-environment relationship can be brought about in several ways, and it may be convenient to distinguish the three types of situation. The first occurs when the organism is overwhelmed by an external stimulus for which it has no adequate adjustive response available and from which it has no means of immediately escaping; this may happen, for instance, in the case of a stimulus too intense for the capacities of the organism, such as electric shock; such situations may be characterized as *traumatic*. Second, in a *frustration* situation the adequate object for an aroused drive or expectation essential to the motivational structure of the organism is not forthcoming from the environment; the result is that, instead of extinction of drive-instigated behavior, there is continued internal stimulation which the organism cannot reduce. Third, there are situations of *conflict*, where the simultaneous arousal of two equally strong drives giving rise to mutually incompatible behavior tendencies makes it impossible for the organism to make effective use of what the environment does offer.

From the point of view of the disruption that occurs in the organism-environment relationship, it is important to note that stress situations are not sharply differentiated from nonstressful situations. But whereas minor disruptions leave no aberrant symptoms behind, and on the contrary serve the organism by calling into play new adjustive responses, there is a point in the continuum beyond which this ceases to be true; beyond this point the responses evoked begin to show those nonadaptive characteristics that are found whenever an organism is caught between its urge to find a means of adjusting and its complete inability to do so. We shall now turn to an examination of these characteristics before attempting to reach an understanding of the mechanisms responsible for their appearance.

Characteristics of Behavior under Stress

The behavioral symptoms which emerge under stress refer, first, to changes in general activity and, second, to changes in the learning process.

Changes in general activity can be classified under two dimensions of behavior: the *rate* and the *range* of activity. The former refers to the excitatory-inhibitory classification which Pavlov proposed and which entails the shift of activity toward either one of the two extremes of this dimension when stress is applied. In the excitatory type, which tends to be more common, behavior is greatly speeded up and disorganized, and the fine adjustment of reaction to stimuli found under normal circumstances is lost. Muscular tremor, vocalization, disturbances in respiration and pulse, mydriasis, diminished control over micturition and defecation, changes in blood pressure, the loss of previously established conditioned reflexes, and a sensitization phenomenon in which the animal is "triggered off" by the least stimulus—all these are symptoms of the excitatory type. In the inhibitory type, on the other hand, disorganization of behavior assumes a different form, for here the shift of activity is toward the other extreme, and a decreased degree of sensitivity and a slowing down of motor responses are accordingly found. Activity may even cease altogether for long periods, and a cataleptic-like immobility manifests itself in the animal.

The second dimension of behavior with which we are here concerned, the range of activity, refers to the general constriction of functioning which occurs in a stress situation. An organism not under stress will show a fairly wide and varied range of activity in a problem situation before such trial-and-error behavior becomes narrowed down and directed toward the set of responses most likely to prove adaptive. Under stress, however, no such directed variability is apparent, and behavior, as Hamilton and Krechevsky (16) have shown experimentally, tends to lose its plasticity and instead assumes a marked stereotypy. Failure to benefit from previous experience and the tendency to persist within a narrow range of responses are thus further characteristics of behavior under stress.

Turning to the learning process under stress, we find that there are certain respects in which this process differs from the kind of learning found in nonstressful situations. These differences manifest themselves (*a*) in the increased rate of acquisition of certain responses, (*b*) in their persistence in a stereotyped form for long periods without reinforcement, and (*c*) in their unadaptive character. In fact, all three of these characteristics may be said to be aspects of the same feature, namely, the greatly increased sensitivity of the learning mechanism operative under stress, which fixates whatever response is dominant at the time and prevents its being extinguished even when it is followed by nothing but unfavorable consequences. The ease of conditioning of certain pathological responses to stimuli associated with the stress experience has been commented on, *inter alia*, by Gantt (12) and Lichtenstein (21); and the stability of such responses and their persistence as stereotyped fixations have been noted in nearly all experimental studies on stress.

The third characteristic of the learning process operative under stress, namely, the unadaptive nature of the responses that are fixated, is shown in the elimination of the trial-and-error behavior usually found in problem-solving situations. Instead of a search taking place for the response that would be most effective in adjusting to the situation, it appears that whatever behavior pattern happens to be dominant at the time stress is applied will become fixated. For instance, Ullman (33), by giving shock to rats at the moment of feeding, was able to produce a compulsive eating symptom in the animals; Wolpe (36) noticed that a cat which happened to be micturating at the time when it was put under stress henceforth always micturated the moment it was again put into the stress environment; and Brown and Jacobs (6) found in the case of rats to whom shock was administered that "fear acts to intensify whatever response is dominant at the moment." Thus, the response pattern that becomes fixated is not necessarily the most efficient that the animal has available at the time of fixation: some of Masserman's (27) cats, for example, chose a response which exposed them to the noxious stimulus (an air blast) rather than one which protected them from it. Only in those cases where stress is introduced more gradually, where its application does not take such a severe and sudden form that the disruption of the organism-environment relationship immediately becomes relatively complete, will the organism first show directed trial-and-error behavior aimed at relieving its plight. But when prolonged inability to bring about adjustment causes the aroused excitation to continue mounting, the disruption of the relationship with the environment will eventually become so great that this problem-solving situation gives way to a stress situation; thereupon the same stereotyped, highly fixative characteristics that we have already noted to be typical of stress will overtake the learning function.

It appears, then, that the learning mechanism operative in stress situations will "freeze" whatever behavior pattern is dominant at the time. Thus the door is opened to all sorts of "irrational" and ill-adapted responses, which are retained in a highly stereotyped form, despite lack of reinforcement from the environment, for much longer periods than are the more variable responses acquired in nonstressful situations.

A Neurophysiological Hypothesis

When seeking an explanation for all these characteristics of behavior under stress, we find not only that few attempts have been made in this direction but that none of the proposals put forward so far has attempted to explain both the changes in general activity and those in the learning process. Pavlov (31) sought to account for the former by his theory of the respective dominance of excitatory or inhibitory cortical processes, but in the absence of empirical confirmation this theory has now been abandoned. More detailed attention has been given to the learning process and the problem of abnormal fixations; in particular, Mowrer's (29) anxiety-reduction theory has been applied to these phenomena by several authors (e.g., 6, 21, 33). An attempt is made by these writers, with the help of the concept of secondary reinforcement, to force learning under stress into the same conceptual framework as learning under normal conditions, and thus to account for both in the light of the law of effect. Eglash (10) has already effectively criticized this attempt; briefly, such a theory has yet to explain why, in the first place, the dominant rather than the most effective response is adopted under stress, often without a preliminary trial-and-error period; and secondly why this response, unlike other drive-reducing responses, assumes such a rigidly stereotyped, unvarying form, which will not even be altered when the animal perceives that the original stress situation is losing its noxious quality. Maier (24), on the other hand, makes no attempt to apply the same laws of learning to "frustration-instigated" responses as to problem-solving behavior; but he does not go beyond stating the distinction, and thus fails to point out a mechanism to account for the peculiarities of learning under stress.

In view of the inadequacy of existing theories, we are forced to look for an alternative explanation, and this, it is here suggested, we may find by turning to the neurophysiological background of the phenomena we have been discussing, with particular reference to the functional relationship of the cortex to the lower cerebral centers. It is generally agreed that in the mature animal under nonstressful conditions this relationship is mainly one of dominance of the cortex over the more primitive centers, so that the activity of the organism is on the whole cortically influenced and modified. Under conditions calling forth emotional behavior, however, mechanisms known to be subcortically situated become activated, and it appears that then the relationship changes and a shift in emphasis occurs from cortex to subcortical centers; or, as Darrow (9) has put it, whenever the cortex is no longer able to deal with afferent excitation, a process of *relative functional decortication* takes place. It may be that the physiological activity of the cortex then becomes inhibited by disruptive discharges from subcortical areas, with the result that the cortex can no longer exert its controlling influence over lower centers. Now this notion, when systematically extended to the whole area of behavior under stress, may well serve as an explanatory scheme for the peculiarities characterizing such behavior. We therefore propose the hypothesis that both in general activity and in learning under stress we are confronted with functions that must be seen primarily in relation to subcortical rather than cortical processes; that stress in fact brings about a shift in dominance from cortical to subcortical centers; and that we must therefore expect differences in behavior corresponding to this change in control. The pathological state resulting from stress may then be regarded as a chronic disturbance in the relationship of the cerebral centers, and the symptoms of this state as a function of such a disturbance.

This hypothesis is made plausible by the existence of certain similarities between the behavior of decorticate preparations and that of intact animals under stress; we shall now turn to these similarities in the belief that they will show the hypothesis to be capable of providing an explanatory scheme for the behavioral phenomena with which we are concerned. The precise neural mechanisms by which a shift in balance between cortex and subcortical centers may be brought about under stress is as yet not clearly understood, and in any case will not concern us here.

Subcortical Control of General Activity

It has long been known that decorticate preparations, in which the forebrain has been removed back to the level of the hypothalamus, will manifest certain marked changes in behavior related mainly to a considerable lowering of the emotional threshold. But it is only recently that the cruder methods of ablation have been replaced by more accurately placed lesions, so that, as in the work of Bard and Mountcastle (2) and of Spiegel, Miller, and Oppenheimer (32), the neural mechanisms necessary to the expression and suppression of "sham rage" can be somewhat more narrowly defined. It appears from these studies that the typical decorticate pattern of behavior will not be brought about as long as the rhinencephalon is left intact, but that once lesions impinge upon this structure, and particularly upon the amygdaloid nuclei, interrupting the tracts from the archicortex to the posterior hypothalamus, the behavior of the organism changes and various symptoms, presumably release phenomena, will manifest themselves. Thus, the degree of emotional sensitivity to stimulation of the operated animal is greatly heightened; even the slightest stimulus will elicit all the signs of violent excitation: biting, trembling, struggling, vocalization, piloerection, mydriasis, high pulse rate and blood pressure, and disturbance in respiration. Similarly, the decreased control over micturition and defecation appears to be due to the release of these functions from higher cerebral control. If in addition the neocortex is affected by lesions, a stereotyped, crude, and constricted form of behavior appears instead of the directed variability of activity shown by normal animals. The work of Klüver and Bucy (19) on monkeys appears to contradict these findings, for similar lesions brought about markedly placid behavior; but, as Lindsley (22) suggests, the difference may lie in the fact that the temporal lobes received somewhat different treatment, or it may be that the inhibitory and excitatory functions which in carnivores lie in the rhinencephalon have migrated farther rostrally in primates. All the symptoms denoting released emotional excitability are, as we have seen, also typical of the changes in the rate and range of activity shown by animals of the excitatory type under stress, and a comparison of these two classes of behavior suggests indeed that they may both stem from the same neurophysiological basis. This would support our hypothesis that under stress the organism is no longer functioning under the same neural control as in a nonstressful situation, but rather that activity is predominantly controlled at subcortical levels.

There is, however, one important difference between the two conditions: emotional after-discharge is shown by organisms under stress but not by the operated animals. Whereas the latter cease their excitement the moment the arousing stimulus is removed, animals that have been under stress may continue in an agitated state for long periods after the experience. It has been suggested by various writers that such self-sustained activity is due to the establishment of thalamocortical reverberating circuits; if this is the case, it would show that, even though much of behavior under stress can be understood as being due to a restriction of cortical functioning and a release of subcortical centers, the cortex is nevertheless continuing to play a part, namely, in the maintenance and support of the aroused excitation.

When we come to consider the inhibitory type of reaction, which, as we have seen, is also shown in some cases under stress, the comparison with the decorticate pattern breaks down. It is likely that we are here confronted with a different kind of disturbance of the relationship of the cortex to the lower cerebral centers, but so far no evidence has come to hand which enables us to gain more precise insight into the nature of this disturbance.

Learning at Subcortical Levels

We have seen that the learning process under stress tends to have characteristics that are different from those operating under nonstressful conditions. If the hypothesis is to be borne out that under stress lower centers take over the control of behavior, we are required to find the mechanism responsible for these characteristics at a neural level below that of the cortex. Is there any evidence to suggest the existence of such a mechanism?

At one time it was thought that learning is an exclusively cortical function, and Pavlov (31), in particular, vigorously opposed all suggestions that the neural locus of conditioning could be subcortical. This view, however, has since been challenged. Though our knowledge of subcortical learning is still meager, several sources of evidence indicate not only that learning is in fact possible subcortically but also that for a certain form of learning the sensitivity of the subcortical mechanisms, when freed from higher inhibition, is very much greater than that of cortical learning mechanisms.

Our first line of evidence comes from the work of Culler and Mettler (8), who have shown that conditioning to shock normally entails two stages. In the first stage a generalized and diffuse pattern of skeletal behavior becomes evident, while in the second a more precise and adaptive response is selected. The first stage (which is the duplicate of the unconditioned response to shock) occurs within a surprisingly short time; the second stage is more gradual and results in the emergence of a localized, precisely adapted response, which is adopted by the animal as the most economic and efficient way of solving the particular problem confronting it. Culler and Mettler have found that the decorticate animal is capable of the first but not of the second stage; that is, it is just as speedily conditionable in a random, diffuse way as intact animals but is incapable of adopting a precise and efficient response. This suggests that the first stage utilizes subcortical learning mechanisms whereas the second requires cortical mechanisms. (It is this later stage which Pavlov had in mind when he wrote of the difficulty of obtaining conditioning at subcortical levels.) It must therefore be concluded that subcortical centers are highly sensitive to "simple diffuse conditioning (direct discharge of substitute impulses into existing action-systems)" (8, p. 301), but that selection of the most adaptive response to fit the problem can only take place when cortical processes are operating.

These conclusions are confirmed by experiments on conditioning under curare and erythroidine—drugs which have been said to suppress cortical activity and thus enable subcortical processes to function independently. Girden (15) has clearly shown that under such conditions "the rapid development and prolonged retention of the drug-state CR" are particularly striking, and that this applies to both skeletal and blood pressure CR's. Once again the learning that is observed proceeds along simple associative lines and does not represent the more refined, problem-solving type of learning.

However, in a further investigation Bromiley (5) obtained different results; his decorticate dog showed the usual type of learning seen in intact animals. Also using an instrumental conditioning procedure, Bromiley found the animal to be capable of giving adaptive responses which, according to the acquisition scores, were not learned at a notably speedy rate. Just what factor accounts for this difference in findings remains at present a puzzle, though one possibility lies in the fact that Bromiley's operation was almost wholly confined to the neocortex, whereas Culler and Mettler also brought about considerable destruction in the archicortex. Only much needed further research in this area will, however, give us the ultimate explanation.

But in the meantime we may note that other data on subcortical learning do support our hypothesis. Zélény and Kadykov (37) have reported their findings on the learning function of a dog after cortical extirpation. A conditioned response established to a certain tone appeared also to neighboring tones; despite 300 presentations of one of these latter tones without reinforcement, extinction failed to occur, though there was some discrimination in the form of a weakening of the response without affecting the strength of the CR to the original tone. These authors also noted the speed with which an olfactory CR developed in their dog, as only seven combinations were necessary for its acquisition.

Another source of evidence on subcortical learning comes from the work of Gantt (12) on the cardiac rate in normal animals. Gantt found that the cardiac component of a conditioned response is a much more sensitive index of learning than are the skeletal components of the reaction, that the former is more quickly formed and persists much longer than the latter. Even two years after the secretory and skeletal parts of a food response had become extinguished, the cardiac component was still retained. A similar finding was made by Anderson and Parmenter (1), who noticed that a sheep, which had been conditioned first to a 10-second and then to a 20-second interval, persisted in showing cardiac acceleration every 10 seconds long after the motor response had been retrained to the longer interval. This phenomenon, Gantt has pointed out, reveals that a basic dysfunction may exist in a normal organism, "a kind of cleavage between the outer specific muscular and the inner cardiac expression" (13, p. 50). The subcortically controlled cardiac reaction is fixated more readily and can persist without reinforcement much longer than the cortically controlled specific skeletal response.

Finally, Hebb (17) has drawn attention to a kind of learning established early in infancy which has all the properties of a reflex in that it is little affected by set and is highly resistant to extinction. As an example of such responses he mentions the eyeblink to a rapidly approaching object, which has been shown to be learned in infancy and to be independent of reinforcement. From the anatomical studies of Conel (7) we may surmise that in early infancy subcortical but not cortical centers are functioning; consequently there can be little doubt that we are confronted here with yet another instance of subcortical learning.

It appears from these observations that there is some evidence suggesting that subcortical centers are indeed highly sensitive learning mechanisms, but that this applies only to simple associative learning. A number of studies, reviewed by Morgan (28), have shown that some discriminative learning involving rather simple habits is possible subcortically, but acquisition and extinction scores show that this type of situation does not yield the phenomena of sensitivity that we have found for associative learning. In the case of skeletal responses, increased sensitivity is probably a release phenomenon, which is suppressed when the cortex is functioning and comes into evidence only as the result of cortical ablation. The cardiac rate, on the other hand, appears to be sensitive even in the presence of the cortex and is therefore presumably not subject to cortical inhibition to the same degree as skeletal responses—a conclusion to which Gellhorn (14) has also given his support.

Whether or not the release phenomenon which we have noted can be brought about by lesions identical to those releasing general activity remains to be tested. Experiments on subcortical learning following partial decortication have so far yielded diverse results. Wing and Smith (35) found extinction of a CR to light to proceed much more slowly after the removal of the striate area than before, but in a subsequent publication Wing (34) has questioned the validity of this conclusion. Marquis and Hilgard (26) found that the conditioned eyelid response to light, which they showed to be subcortically acquired, failed to extinguish in monkeys whose occipital lobes had been removed; however, they also made this observation in normal monkeys but not in normal or operated dogs (25). No conclusions, either for or against our hypothesis, can therefore be drawn from this area.

On the other hand, we have seen that in those instances in which the cortex is indubitably no longer exerting an inhibitory influence on subcortical centers, the characteristics of learning under stress may also appear at a subcortical level. The evidence does therefore appear to be consistent with the view that in situations of stress the cortex ceases to be the chief controlling and integrating center, and the responses then acquired are the result of subcortical learning. One might indeed view the cortex as essentially an organ of adaptation, which prevents the acquisition of completely nonadjustive responses and brings about extinction as soon as any response is no longer reinforced. In its absence, however, more primitive and less adaptive mechanisms come into force, and it is these mechanisms that may be said to account for the phenomena found under conditions of stress. Thus it appears that the subcortical centers, owing to their highly sensitive nature, fixate whatever pattern of behavior happens to be contiguous at the time, and that this pattern is then reactivated by all further appearances of the stressful stimulus, as well as by any previously neutral stimulus occurring in the same environment and thereafter also firmly associated with the response.

It will be seen that, although the present theory agrees with Maier's (24) frustration theory in postulating that fixations cannot be explained by the same laws of learning that govern "goal-oriented" responses, it goes further. For while Maier rightly protests against attempts to fit these pathological responses into the Procrustean bed of problem-solving behavior, he makes no attempt to explain the mechanism by which their "freezing" comes about. This mechanism, it is here suggested, can be found in the special nature of subcortical learning.

Conclusions

The functional relationship between cortex and lower cerebral centers is an area about which our knowledge is still scanty, and the experimental data with which we are provided are as yet limited. But what evidence there is suggests that in any attempt to understand behavior it is fruitful to take this neurophysiological relationship into account as an intervening variable and to relate the changes to which it is subject to behavioral phenomena. This the present paper has attempted to do, and though the hypothesis of subcortical dominance under stress still stands in need of direct empirical testing, it does appear, as we have seen, that there are some grounds for believing this hypothesis to be capable of explaining the characteristics of behavior under stress. It is true that we must beware of taking too extreme a position in this matter, that the cortex is not in fact an isolatable unit physiologically (it may well be that the dichotomy between cortex and subcortical centers is, physiologically speaking, too crude). Also the part the cortex plays in the maintenance and elaboration of excitation subsequent to stress-induced breakdown must not be overlooked. Nevertheless, it seems likely that a relation exists between the extent of disruption of the organism-environment relationship and the relative degree of control exercised over behavior by cortical and subcortical centers. The hypothesized relation is such that whenever aroused excitation is unable to find a readily available response channel through which to discharge, cortical control becomes weakened and the more primitive action systems controlled by the diencephalon become correspondingly prominent. In problem-solving situations where disruption of the organism-environment relationship is not severe, the subcortical element is probably relatively small; it merely adds an emotional component to behavior that otherwise still bears the hallmarks of cortical variability and direction. But once the disruption becomes extreme, as under stress, the shift in dominance is likely to be more drastic, and behavior will then change accordingly. The possibility that such a change takes place is especially relevant when we seek to explain the learning process under stress, for any approach which insists, as the anxiety-reduction theory does, on fitting the same principles that govern learning in nonstressful situations to learning under stress runs into the danger of making unjustified generalizations. But if an attempt is made to link the two types of behavior to their neural bases, it seems indicated that the reflex-like quality of stress-fixated responses is a function, not of some sophisticated learning process, but of a relatively primitive neural level at which automatic and stereotyped responses are the rule.

The specialized nature of subcortical learning may well be a process that underlies also a variety of other phenomena in which for diverse reasons subcortical centers are likely to be dominant. These include the fixating instead of extinguishing effect which punishment has been found to have under certain circumstances (30); the particularly speedy and permanent type of learning found in certain critical periods of development in various species, which Lorenz (23) has termed "imprinting"; and the enduring nature of some forms of learning in infancy, when, as Bousfield and Orbison (3) have argued, the organism is still in a "precorticate" condition.

Summary

1. A stressful situation may be described as one in which a major disruption of the relation of an organism to its environment has taken place; it is brought about when a highly motivated organism is unable to find an adjustive response to the problem confronting it. This may occur under conditions variously described as trauma, frustration, and conflict.

2. As a result of stress certain changes manifest themselves in general activity and in the learning process. Both the rate and the range of general activity become altered, while the learning process is characterized by a highly increased degree of sensitivity, as shown by the altered rates of acquisition and extinction, and by the tendency to fixate whatever response is dominant at the time.

3. None of the theories so far advanced has proved successful in explaining the mechanisms responsible for all these changes. The hypothesis is here advanced that under stress a shift in emphasis occurs from cortical to subcortical centers, and that consequently behavior under stress must be seen primarily in relation to subcortical processes.

4. This hypothesis is supported by the similarity of general activity under stress to that of decorticate animals.

5. The hypothesis receives further support from the nature of subcortical learning; according to various lines of evidence, such learning is highly sensitive for purely associative learning. This could account for the characteristics of learning under stress, and possibly also for certain other phenomena that have been indicated.

REFERENCES

1. Anderson, O. D., & Parmenter, R. A long-term study of the experimental neurosis in the sheep and the dog. *Psychosom. Med. Monogr.*, 1941, 2, Nos. 3 & 4.

2. Bard, P., & Mountcastle, V. B. Some forebrain mechanisms involved in expression of rage with special reference to suppression of angry behavior. *Ass. Res. nerv. ment. Dis.*, 1947, 27, 362–404.

3. Bousfield, W. A., & Orbison, W. D. Ontogenesis of emotional behavior. *Psychol. Rev.*, 1952, 59, 1–7.

4. Bowlby, J. Maternal care and mental health. *World Hlth Org. Monogr. Ser.*, 1951, No. 2. (United Kingdom: His Majesty's Stationery Office; United States: Columbia Univer. Press.)

5. BROMILEY, R. B. Conditioned responses after removal of the neocortex. *J. comp. physiol. Psychol.*, 1948, **41**, 102–109.

6. BROWN, J. S., & JACOBS, A. Fear in motivation and acquisition. *J. exp. Psychol.*, 1949, **39**, 747–759.

7. CONEL, J. L. *The postnatal development of the human cerebral cortex.* Vol. 1–4. Cambridge: Harvard Univer. Press, 1940–1951.

8. CULLER, E., & METTLER, F. A. Conditioned behavior in a decorticate dog. *J. comp. Psychol.*, 1934, **18**, 291–303.

9. DARROW, C. W. Emotion as relative functional decortication: the role of conflict. *Psychol. Rev.*, 1935, **42**, 566–578.

10. ECLASH, A. The dilemma of fear as a motivating force. *Psychol. Rev.*, 1952, **59**, 376–379.

11. FULLER, J. L. Situational analysis: a classification of organism-field interaction. *Psychol. Rev.*, 1950, **57**, 3–18.

12. GANTT, W. H. *Experimental basis for neurotic behavior.* New York: Paul B. Hoeber, 1944.

13. GANTT, W. H. Psychosexuality in animals. In P. H. Hoch & J. Zubin (Eds.), *Psychosexual development in health and disease.* New York: Grune & Stratton, 1949. Pp. 33–51.

14. GELLHORN, E. *Autonomic regulations.* New York: Interscience Publishers, 1943.

15. GIRDEN, E. The dissociation of blood pressure conditioned responses under erythroidine. *J. exp. Psychol.*, 1942, **31**, 219–231.

16. HAMILTON, J. A., & KRECHEVSKY, I. Studies in the effect of shock upon behavior plasticity in the rat. *J. comp. Psychol.*, 1933, **16**, 237–253.

17. HEBB, D. O. *The organization of behavior.* New York: Wiley, 1949.

18. KARDINER, A. *The traumatic neuroses of war.* New York: Paul B. Hoeber, 1941.

19. KLÜVER, H., & BUCY, P. C. Preliminary analysis of functions of the temporal lobes in monkeys. *Arch. Neurol. Psychiat., Chicago*, 1939, **42**, 979–1000.

20. LAZARUS, R. S., DEESE, J., & OSLER, SONIA F. The effects of psychological stress upon performance. *Psychol. Bull.*, 1952, **49**, 293–317.

21. LICHTENSTEIN, P. E. Studies of anxiety: I. The production of a feeding inhibition in dogs. *J. comp. physiol. Psychol.*, 1950, **43**, 16–29.

22. LINDSLEY, D. Emotion. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 473–516.

23. LORENZ, K. The companion in the bird's world. *Auk*, 1937, **54**, 245–273.

24. MAIER, N. R. F. *Frustration: the study of behavior without a goal.* New York: McGraw-Hill, 1949.

25. MARQUIS, D. G., & HILGARD, E. R. Conditioned lid responses to light in dogs after removal of the visual cortex. *J. comp. Psychol.*, 1936, **22**, 157–178.

26. MARQUIS, D. G., & HILGARD, E. R. Conditioned responses to light in monkeys after removal of the occipital lobes. *Brain*, 1937, **60**, 1–12.

27. MASSERMAN, J. H. *Behavior and neurosis.* Chicago: Univer. of Chicago Press, 1943.

28. MORGAN, C. T. The psychophysiology of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 758–788.

29. MOWRER, O. H. *Learning theory and personality dynamics.* New York: Ronald, 1950.

30. MUENZINGER, K. F. Motivation in learning. I. Electric shock for correct response in the visual discrimination habit. *J. comp. Psychol.*, 1934, **17**, 267–277.

31. PAVLOV, I. P. *Lectures on conditioned reflexes.* New York: International Publishers, 1928.

32. SPIEGEL, E. A., MILLER, H. R., & OPPENHEIMER, M. J. Forebrain and rage reaction. *J. Neurophysiol.*, 1940, **3**, 338–348.

33. ULLMAN, A. D. Compulsive eating symptoms in rats. *J. comp. physiol. Psychol.*, 1951, **44**, 575–581.

34. WING, K. G. The role of the optic cortex of the dog in the retention of learned responses to light: conditioning with light and shock. *Amer. J. Psychol.*, 1946, **59**, 583–612.

35. WING, K. W., & SMITH, K. U. The role of the optic cortex in the dog in the determination of functional properties of conditioned reactions to light. *J. exp. Psychol.*, 1942, **31**, 478–496.

36. WOLPE, J. Experimental neuroses as learned behavior. *Brit. J. Psychol.*, 1952, **43**, 243–268.

37. ZÉLÉNY, G. P., & KADYKOV, B. I. [Contribution to the study of conditioned reflexes in the dog after cortical extirpation.] *Méd. exp. Kharkov*, 1938, No. 3, 31–34. (*Psychol. Abstr.*, 1938, **12**, No. 5829.)

(Received December 4, 1953)

Psychological Review
Vol. 61, No. 5, 1954

A NOTE ON THE CIRCULAR RESPONSE HYPOTHESIS

WAYNE DENNIS

Brooklyn College

Perhaps no visual aid has been more widely reproduced in psychological textbooks than a diagram by F. H. Allport which appears in his *Social Psychology* (1924). It illustrates the fact that when a child speaks his voice normally stimulates his own ears. Allport proposes that by virtue of this fact an association is formed between the hearing of a sound and the utterance of this sound by the child. This principle, usually called the circular response hypothesis, is employed to account for self-imitation (repetition of one's own acts), imitation of others, and sympathetic responses. The principle is not limited to vocal responses but applies wherever a response causes stimulation of the responding organism. The circular response hypothesis is frequently assigned a significant role in social psychology and child psychology, as well as in the introductory course. Among the recent textbooks which make use of this principle are the following: Boring, Langfeld, and Weld (7, p. 44), Dashiell (12, p. 532), Hurlock (19, p. 206), Sargent (26, p. 228), and Stagner and Karwoski (28, p. 341).

This theory is so common in psychology that authors frequently feel no responsibility for referring to its history. When an origin is mentioned, it is often credited to Allport, apparently because he was responsible for the illustration referred to above. The hypothesis is sometimes attributed to Baldwin (3, 4). Among the authors who credit Allport or Baldwin with the origin of the theory are Curti (11, p. 258), Folsom (13, p. 93), and Merry and Merry (23, p. 79).

The aim of the present paper is to call attention to the fact that the theory is an old one. It is an heirloom derived from our associationistic heritage, redecorated to give it a modern air.

Nowadays the theory is usually stated in terms of conditioning. This fact leads to the impression that the concept arose from conditioned response theory. In fact, the earliest description of the circular response hypothesis is two centuries old. It was clearly stated by Hartley in 1749.

Hartley's application of the principle to infant speech reads as follows:

> I will, in the next place, give a short account of the manner in which we learn to speak . . . Suppose now the muscles of speech to act . . . It is evident that an articulate sound, or one approaching thereto, will sometimes be produced by this conjoint action of the muscles of the trunk, larynx, tongue and lips; and that both the articulate and the inarticulate one will often recur from the recurrence of the same accidental causes. After they have recurred a sufficient number of times, the impression which these sounds, articulate and inarticulate, make upon the ear will become an associated circumstance (for the child always hears himself speak at the same time he exerts the action) sufficient to produce a repetition of them. And thus it is that children repeat the same sounds over and over again for many successions, the impression of the last sound upon the ear exciting a fresh one and so on till the organs be tired (15, p. 109).

He noted also that the same principle provides a basis for verbal imitation:

> It follows, therefore, that if any of the attendants make any of the sounds familiar to the child, he will be excited by this impression, considered as an associated circumstance, to return it (15, p. 109).

A few pages further on, Hartley gave a more general account of the origin of imitation in children:

> They see the actions of their own hands, and hear themselves pronounce, hence the im-

> pressions made by themselves on their own eyes and ears become associated circumstances, and consequently must, in due time, excite to the repetition of the actions. Hence, like impressions made on their eyes and ears by others will have the same effect; or in other words, they will learn to imitate the actions which they see, and the sounds which they hear (p. 111).

In another place Hartley applied his doctrine of association to the origin of sympathy:

> Now this in children seems to be grounded upon such associations as the appearance and idea of any kind of misery which they have experienced . . . because the connection between the adjuncts of pain and the actual infliction of it has not yet been sufficiently broken by experience as in adults (p. 487).

Hartley's immediate successors in the British association school seem not to have written concerning this hypothesis, probably because they were interested only in the association of ideas and were not interested in muscular movement. While it is difficult to make sure that one has not overlooked some statement in the extensive works of the associationists, we have been able to find references to circular reactions only in Brown (10) and in Bain (2).

The reference in Thomas Brown, whose work was first published in 1820, concerns only sympathy and is not very explicit. Brown said:

> Many of the phenomena of sympathy, I have little doubt, are referable to the laws to which we have traced the common phenomena of suggestion or association. It may be considered as a necessary consequence of these very laws, that the sight of any of the common symbols of internal feeling, should recall to us the feeling itself . . . (10, p. 106).

No doubt some of the "common symbols of internal feelings" referred to by Brown are responses made by a person in distress which the person himself can sense, such as crying, shrinking, and trembling. If Brown intended to include such "symbols," then he stated the circular hypothesis, but only with special reference to sympathy.

Bain, whose *Senses and the Intellect* first appeared in 1855, was specifically concerned with the circular response hypothesis in connection with the development of verbal imitation. Bain stated:

> The sound spoken is also heard; besides the vocal exertion there is a coincident impression on the ear; an association grows up between the exertion and the sensation, and, after a sufficient time, the one is able to recall the other. The sensation, anyhow occurring, brings on the exertion; and when by some other person's repeating the syllable, the familiar sound is heard, the corresponding vocal act will follow (2, p. 415).

The wording here is reminiscent of Hartley. Bain was, of course, familiar with this writer, but possibly had overlooked or forgotten Hartley's discussion of what is here called the circular response hypothesis. Since Bain made no reference to Hartley or to anyone else in this connection, it may be that he developed this idea independently.

In surveying the history of the hypothesis under consideration, we come next to Baldwin. Prior to Baldwin the concept under discussion had received no name. Baldwin originated the term "circular response" and the term has stuck. However, "circular response" as used by Baldwin did not have the limited meaning that it has at the present time. Baldwin at times used it in its present significance, but in addition he used it to refer to other very diverse phenomena.

Among Baldwin's statements is the following: "The child who has learned to make a sound, then makes it by association whenever he hears it" (3, p. 284). Other such statements can be found in Baldwin's writings. He denied that this was an original proposal, saying: "I know that this is a widespread view" (3, p. 284). It is interesting to note that whereas he stated this view, he criticized the notion that this type of association is sufficient to account for the learning of language, stating that in order to learn language the child must develop a tendency to imitate *all* sounds (3, p. 284), not merely a tendency to imitate those which are already in his repertoire.

We turn next to Stout. In discussing the child's imitation, Stout in 1903 (30, p. 82) made the brief statement that imitation presupposes a motor association between the perception of the act to be imitated and the more or less similar movements that the child has already learned to perform. "Hence, the more he has already learned to do, the more he can do in the way of imitation . . ." Stout (30, p. 158) stated that such associations are the basis for verbal imitations. He gave no reference to earlier enunciations of these views.

French (14), writing in the year following the appearance of Stout's book, indicated that he was stimulated by Stout's brief comments to work out a fuller account of imitation along the same lines. This he did in a very clear-cut manner. He explained sympathy by the same mechanism. French referred to Baldwin as well as to Stout.

We have not seen the 1913 edition of Bechterev's *Objective Psychology*, but Lewis (21) credits this book with an expression of the circular response hypothesis. The principle is clearly stated in the 1926 English edition of Bechterev's *General Principles of Human Reflexology* (5, p. 209).

H. C. Brown (9) stated the hypothesis briefly in 1916; he credited Bechterev with the idea.

By 1920, conditioning principles were well-known in America, and were widely applied as explanations in various fields. Between 1920 and 1930 several instances of the derivation of the circular response hypothesis from conditioning concepts can be cited. Most of these seem to have been independently derived.

First among the recent rediscoverers was Humphrey (17, 18). His formal statement is as follows: "Imitative action may be defined as action involving a conditioned reflex, the secondary stimulus of which is similar to the reaction" (17, p. 5). He indicates that the child's own responses are sufficient to set up such conditioning. Humphrey was acquainted with Baldwin's concept of "circular reaction" and proceeded to show that it is not identical with the conditioning theory. This is quite true, as we have shown above, but Humphrey seems not to have noticed that although Baldwin sometimes used the term "circular response" to mean something different from what it means today, nevertheless Baldwin was acquainted with the idea that responses may become associated with their own sensory effects, and, in fact, spoke of this idea as a "wide-spread view."

The first appearance of the circular response hypothesis in an American textbook occurred in Smith and Guthrie's *General Psychology in Terms of Behavior*, which was first published in 1921. These authors wrote:

> Practically all imitative behavior is made up of conditioned responses. . . . The dependence of imitation on learning is well illustrated by language acquisition . . . sounds . . . accompany the movements that produce them and, because the vowels are sustained and the consonants either sustained or repeated, these sounds also precede the movements that continue or iterate them. They thus become the conditioning stimuli for their own production, so that when uttered by others they are imitated by the baby (27, p. 132).

Smith and Guthrie did not indicate whether or not they were acquainted with presentations of this idea on the part of others.

The first German edition of Koffka's *Growth of the Mind* also appeared in

1921. Koffka referred to the circular response theory of imitation and credited it to Baldwin (20, p. 310).

McDougall, in 1923, gave a concise account (22) of vocal imitation in terms of association. Since he did so without any citation of previous writers, we are left in doubt as to whether he considered that this hypothesis was a matter of common knowledge among psychologists or whether he believed himself to be making an original proposal.

The second edition of Stern's *Psychology of Early Childhood* appeared in German in 1923 and in English in 1924. In this edition Stern (29, p. 91) described the circular response theory of imitation and ascribed it to Baldwin.

As we indicated earlier, Allport (1) expressed the concept of the circular response very completely in his *Social Psychology* published in 1924. He applied it to language acquisition, to imitation, and to sympathy. Allport states that he arrived at this theory independently, but before his book was published he came across the comparable statement by Smith and Guthrie. He does not appear to have known of other origins.

It is clear that Holt's discussion (16) of circular responses, which has attracted much attention, had an extensive historical background. Of this, Holt, the philosopher-psychologist, was certainly somewhat aware. He gave credit to earlier discussions by Baldwin (3), Bok (6), and Humphrey (18), but he did not mention Hartley, Brown, or Bain. Holt's use of this principle is much more extensive and thoroughgoing than that of any of his predecessors, and his ambitious attempt to use this and other concepts to show that all of human behavior is learned, whether it proves to be convincing or not, deserves the attention which it has received.

The foregoing account is sufficient to indicate that the circular response hypothesis of iteration, imitation, and sympathy has a long history, and that it has probably had several independent origins. It remains to be noted that it has never been subjected to an experimental test. Can it be tested? We are not sure. Perhaps, in the hands of some ingenious experimenter, it can become a testable hypothesis. But its presence in psychology is due, not to research, but to the recurrence of associationistic thought in psychology. It is interesting that an arm chair hypothesis, two hundred years old, which has never been tested nevertheless continues to find a secure place in our textbooks. Perhaps its recent affiliation with C-R theory has tended to give it respectability. It is likely, too, that it continues not only because it has been recast in C-R terms but also because no rival theory has arisen to contest its place. At any rate, it provides an interesting example of the perseveration of unsupported theory in a field which prides itself upon its empiricism.

To summarize, many authors have made use of the hypothesis that iteration, imitation, and sympathy arise because a response necessarily becomes associated with its own sensory consequences. We have shown that priority for this hypothesis belongs to Hartley, who clearly stated it in 1749. The same hypothesis seems to have been developed independently by several other writers, including Bain, Humphrey, Smith and Guthrie, McDougall, and Allport. Although this hypothesis is widely accepted, no experimental test of it seems ever to have been attempted. The facts just surveyed seem to justify the conclusions that psychologists, even eminent ones, have been poorly informed concerning their predecessors' treatments of the circular response hypothesis; and that an untested theory, which has plausibility to support it, can still find a place in psychological textbooks.

REFERENCES

1. ALLPORT, F. H. *Social psychology.* New York: Houghton Mifflin, 1924.
2. BAIN, A. *The senses and the intellect.* (3rd Ed.) New York: D. Appleton & Co., 1872.
3. BALDWIN, J. M. *Mental development in the child and the race.* (2nd Ed.) New York: Macmillan, 1897.
4. BALDWIN, J. M. *Social and ethical interpretations in mental development.* (3rd Ed. Rev.) New York: Macmillan, 1902.
5. BECHTEREV, V. M. *General principles of human reflexology.* (Trans. from 4th Ed.) London: Jarrolds, 1928.
6. BOK, S. T. The development of reflexes and reflex tracts. *Psychiat. en Neurol. Bladen,* 1917, 21, 281–303.
7. BORING, E. G., LANGFELD, H. S., & WELD, H. P. *Foundations of psychology.* New York: Wiley, 1948.
8. BRITT, S. H. *Social psychology of modern life.* New York: Farrar & Rinehart, 1941.
9. BROWN, H. C. Language and the associative reflex. *J. Phil. Psychol. sci. Method,* 1916, 13, 645–649.
10. BROWN, T. *Lectures on the philosophy of the human mind.* Hallowell: Masters, Smith & Co., 1850.
11. CURTI, MARGARET W. *Child psychology.* (2nd Ed.) New York: Longmans, Green, 1938.
12. DASHIELL, J. F. *Fundamentals of general psychology.* (3rd Ed.) New York: Houghton Mifflin, 1949.
13. FOLSOM, J. K. *Social psychology.* New York: Harper, 1931.
14. FRENCH, F. C. The mechanism of imitation. *Psychol. Rev.,* 1904, **11**, 138–142.
15. HARTLEY, D. *Observations on man, his frame, his duty and his expectations.* (5th Ed.) London: Richard Curttwell, 1810. (First edition, 1749.)
16. HOLT, E. B. *Animal drive and the learning process.* New York: Holt, 1931.
17. HUMPHREY, G. Imitation and the conditioned reflex. *Ped. Sem.,* 1921, **28**, 1–21.
18. HUMPHREY, G. The conditioned reflex and the elementary social reaction. *J. abnorm. soc. Psychol.,* 1922, **17**, 113–120.
19. HURLOCK, ELIZABETH B. *Child development.* (2nd Ed.) New York: McGraw-Hill, 1950.
20. KOFFKA, K. *The growth of the mind.* (Trans.) New York: Harcourt, Brace, 1925.
21. LEWIS, M. M. *Infant speech.* New York: Harcourt, Brace, 1936.
22. McDOUGALL, W. *Outline of psychology.* New York: Scribner's, 1923.
23. MERRY, F. K., & MERRY, R. V. *From infancy to adolescence.* New York: Harper, 1940.
24. MUNN, N. L. *Psychological development.* New York: Houghton Mifflin, 1938.
25. MUNN, N. L. *Psychology.* (2nd Ed.) New York: Houghton Mifflin, 1950.
26. SARGENT, S. S. *Social psychology.* New York: Ronald, 1950.
27. SMITH, S., & GUTHRIE, E. *General psychology in terms of behavior.* New York: Appleton-Century-Crofts, 1921.
28. STAGNER, R., & KARWOSKI, T. F. *Psychology.* New York: McGraw-Hill, 1952.
29. STERN, W. *Psychology of early childhood.* (Trans. from 3rd Ed.) New York: Holt, 1924.
30. STOUT, G. F. *The groundwork of psychology.* New York: Hinds & Noble, 1903.

(Received February 19, 1954)

Psychological Review
Vol. 61, No. 5, 1954

THE SCIENCE OF PERSONALITY: NOMOTHETIC!

H. J. EYSENCK

Institute of Psychiatry, Maudsley Hospital

The cleavage (or perhaps cleft would be a less emotionally charged term) between the nomothetic and idiographic approaches to the study of personality is indeed, as Beck (2) has pointed out in his recent paper on this subject, "a principal and vigorously debated issue before psychology today." It follows that any serious attempt to reconcile and integrate these two opposing views should be given a sympathetic hearing, and should be attentively studied by all concerned with the concept of personality. Careful perusal of Beck's proposals, however, has brought to light what appear to this writer to be a number of fallacies that appear to make his attempt at reconciliation less appealing than it might appear at first sight.

First, let us be clear about the meaning of the words used. As is well known, they were introduced by the philosopher Windelband (14) as yet another set of terms to distinguish the *naturwissenschaftliche* (scientific, nomothetic) way of studying psychology from the *geisteswissenschaftliche* (humanistic, idiographic) manner. Allport (1) was one of the first to bring the concepts into use in Anglo-American psychology, and his exposition clearly indicates the meaning attaching to the words "nomothetic" and "idiographic." "The former [sciences] . . . seek only general laws and employ only those procedures admitted by the exact sciences. . . . The idiographic sciences, such as history, biography, and literature, on the other hand, endeavor to understand some *particular* event in nature or in society." This quotation makes it clear how difficult it would be to reconcile these two points of view; literature, even if called a "science" by Allport (it would be interesting to know the justification for this curious appellation, probably equally repugnant to writers as to scientists), does not lie down easily with psychometrics. Beck has cut the Gordian knot by disowning "idiography" completely, and by rechristening a part of the nomothetic field "idiographic." A brief quotation from his paper will substantiate this argument.

> . . . let it be noted that, so far as concerns the basic procedures of scientific method, the two methods have everything in common. They both have recourse to observation and to experiment. They analyze and resynthesize data. They draw inferences that follow the usual canons of logic, both inductive and deductive. These are the foundational approaches to scientific method (2, p. 253).

This is certainly an appealing picture, but it bears no relation to Windelband's or Allport's definition of these terms. Beck has in effect surrendered the castle of idiographic beliefs; he has given up the basic proposition that idiographic procedures are founded on the view that what he calls the "basic procedures of scientific method" are inapplicable to personality research.

Having thus emptied the term of its usual, and very useful, meaning, he invests it with an entirely new content. Quite arbitrarily, Beck divides the customary type of nomothetic research into two separate steps, one of which he calls nomothetic, the other idiographic. As far as can be deduced from his paper, it would appear that the measurement of isolated traits, such as bravery, or pride, or sense of humor, is to be regarded as nomothetic; it becomes idiographic when we "ask about any person how much bravery does he have, *and* coolness, *and* pride, *and* sense of humor, *and* other variables that fuse into character" (2, p. 253). Nothing here of the complete and total rejection of such nomothetic concepts as traits, which is the main characteristic of the traditional idiographic attitude; instead, we find that when we study traits in combination, we are no longer doing nomothetic research, but idiographic! Having throughout his professional life studied traits in combination, having always paid particular attention to the ways in which they interact, modify each other, and, through their interaction, "[bring] about the total behavior which we identify as a particular personality" (2, p. 254), the present writer notes with surprise that instead of being a hard-bitten nomothetical psychologist, he has in fact always acted on idiographic principles. The reader may recall Molière's *Monsieur Jourdain*, who discovered late in life that he had always been speaking prose!

Bewilderment becomes complete when we hear that factor analysis is recommended as a favorite method of this "new look" idiography. According to Beck: "A universe of traits, variables in mutual interplay, affecting one another, these are the individual. This is the task which the idiographic method undertakes. The specific technique devised to test out the findings in this kind of universe is that associated with Stephenson—the Q technique" (2, p. 358). Beck is apparently referring to the method of intercorrelating persons, introduced by Thomson and Bailes (13), and factor-analyzing the resulting matrix of intercorrelations, introduced by Beebe-Center (3). (Others who have some claim to have introduced this method are Burt [4, 5], Thomson [12], and Stern [11].) This gives us a specific example to illustrate our contention that Beck's "idiography" is nothing but the old-fashioned nomothetic method dressed up in slightly different clothing.

By giving preference to the method of "correlating persons" over the usual method of "correlating tests," and by implying that the former is better suited to the demands of personality research, Beck is clearly adopting the view that these two procedures give different results. It is obvious that if results of two methods are identical, or convertible into one another by some simple mathematical formula, then it is not possible to describe the one as "the specific technique" for testing hypotheses of a certain kind as contrasted with the other. Now Burt (6) and Cattell (7) have discussed this question of convertibility in detail, and there appears to be no doubt that, statistically, factors derived from the intercorrelations between persons (*Q* technique) are transposable from factors derived from intercorrelations between tests (*R* technique). As Cattell (7) points out:

> The belief of some users of *Q* technique that it is fundamentally different from its transpose technique—*R*—and, indeed, a method *sui generis*, has so far been most exhaustively statistically examined and refuted by Sir Cyril Burt. . . . In the writer's experience professional statisticians take the position that there is no doubt about the transposability of factors from a double-centered score matrix though there may be doubt about the exact relation under other and special conditions. . . . *R* and *Q* techniques normally . . . (i.e. without double centering) have the completeness of their transposability slightly restricted by some inevitable mutual losses of information. The losses which then occur are (*a*) of the variance of the first factor (or in some conditions the first two) and (*b*) of the specific factors . . . (7, pp. 506–507).

The rest of Cattell's paper should be carefully studied to enable the relevance of this loss to be evaluated in relation to the question at issue; the conclusion the present writer has come to independently of Cattell's review (cf. 9) agrees completely with Cattell's assess-ment, as well as with that of Burt, in considering the *Q* sort a very questionable procedure from the statistical point of view, which at best simply duplicates factors usually more easily and safely obtained by *R* technique. Beck nowhere answers the far-reaching criticisms made of *Q* technique, nor does he consider the identity of factors produced by analyzing a matrix or its transpose; in view of the practically unanimous verdict of those qualified to judge the statistical issues involved, we must conclude that the method favored by him produces, at best, factors also produced by the arch-nomothetic procedure of correlating tests, while at worst it is beset by so many statistical fallacies as to make results meaningless.

Beck refers to some results obtained by him with the use of this method; he says: "We have succeeded in . . . isolating six schizophrenic reaction patterns. That is, we are describing six patterns within this disease group that differ from one another" (2, p. 358). As he does not give any details, it is not possible to compare his patterns with those found along traditional lines by T. V. Moore (10), or by Wittenborn (15); given comparability of populations used, it may be predicted that there will be considerable similarity. Here again, it is difficult to see precisely what new contribution the *Q* method is supposed to make, or in what way the result is "idiographic"; method and aim alike are the stock in trade of the nomothetic psychologist. Having discussed the issues involved at length, with full experimental documentation, the writer may perhaps be allowed to refer the interested reader elsewhere (9). We may now state our conclusion. Beck has set out to reconcile and integrate the idiographic and nomothetic approaches. Instead of using these terms in their traditional sense, however, he has thrown overboard completely the idiographic conception, and has instead rechristened part of the traditional nomothetic procedure as "idiographic." Renaming different approaches in this arbitrary fashion merely sows seeds of semantic confusion; it does not contribute to the *rapprochement* desired by Beck. The scientific and the literary views of personality are still as different and as opposed to each other as ever, and the only valid conclusion to be drawn from Beck's paper and his implicit withdrawal from the idiographic position is that suggested in the title of this article: *the science of personality must by its very nature be nomothetic.* This is the conclusion to which the writer was led after an extensive examination of the arguments and experiments adduced by many writers in this field (8), and Beck's contribution has strengthened, rather than weakened, belief in the essential correctness of this view.

REFERENCES

1. Allport, G. W. *Personality. A psychological interpretation.* London: Constable, 1938.
2. Beck, S. J. The science of personality: nomothetic or idiographic? *Psychol. Rev.*, 1953, 60, 353–359.
3. Beebe-Center, J. B. *Pleasantness and unpleasantness.* New York: Century, 1933.
4. Burt, C. L. The mental differences between the sexes. *J. exp. Pedag.*, 1912, 1, 273–284.
5. Burt, C. L. *The distribution and relations of educational abilities.* London: P. S. King, 1917.
6. Burt, C. L. Correlations between persons. *Brit. J. Psychol.*, 1937, 28, 59–96.
7. Cattell, R. B. The three basic factor-analytic research designs—their interrelations and derivatives. *Psychol. Bull.*, 1952, 49, 499–520.
8. Eysenck, H. J. *The scientific study of personality.* London: Routledge & Kegan Paul, 1952.

9. EYSENCK, H. J. *The structure of human personality.* London: Methuen, 1953.

10. MOORE, T. V. The essential psychoses and their fundamental syndromes. *Stud. Psychol. Psychiat.*, 1933, 3, 128.

11. STERN, W. *Differentielle psychologie.* Leipzig: Barth, 1911.

12. THOMSON, G. H. *The factorial analysis of human ability.* London: Univer. of London Press, 1948.

13. THOMSON, G. H., & BAILES, S. The reliability of essay marks. *For. Educ.*, 1926, 4, 85–91.

14. WINDELBAND, W. *Geschichte und Naturwissenschaft* (3rd Ed.) Strasburg: Heitz, 1904.

15. WITTENBORN, J. R. Symptom patterns in a group of mental hospital patients. *J. consult. Psychol.*, 1951, 15, 290–302.

(Received January 8, 1954)

Psychological Review
Vol. 61, No. 5, 1954

SIDESTEPS TOWARD A NONSPECIAL THEORY [1]

EDGAR F. BORGATTA
Harvard University

Occasionally man has seen himself in a mirror, and not recognizing himself, has criticized himself before he understood what he was doing.[2] As a result, some information has been validated in part, frequently in contradiction to the beliefs, moral and religious, of his community. While the good books have stated that man should love his neighbor and his brother, the nascent social scientists have been accumulating data concerning *what man is and what his actions are*. This eking out of information, while replete with error, has continued and appears finally to be approaching a scientific footing.

Some studies of man merely record his follies, and blandly state that the road to survival is in their removal. Treatises on social problems, social conflicts, politics, etc. are often no better. Those attempts in the study of man which will prove most useful are those which reach further and further back. Volumes which name maladies and conditions are of little help. We are certain that men are not perfect, and that terms such as neurotic personality or sick society are apt, but these namings should not lull us into feeling we understand man better.[3] It is a simple statement to make: "If all men were better, this would be a better world." And yet, it is a ridiculous one. If all men were better, who knows what it would mean, and what was better? At the same time, however, these very writers do a great scientific service as they focus on certain *special* problems of developmental or of historical circumstances. Thus, Erich Fromm (2) has focused on the changes or sources of security in the transition of extended family to small family, of closed community to open community. Such a special focus aids in the understanding of some current variation, but more fundamental questions remain unanswered. Similarly, the developing theory (which is again a special theory focusing on the contemporary situation) of vertical, diagonal, and sideways mobility, while useful, leaves the important questions untouched. Again, other attempts have been made which are of a *nonspecial* nature, but these have been frequently tangential to the development of explanatory concepts; in particular, they have dealt with description of system, the establishment of frames of reference, or the specification of system-model-structural-functional frameworks.[4]

[1] Colleagues and friends have contributed in so many ways to the development of this theory that it is not possible to credit them individually. *Sidesteps* was added to the original title because it was felt that where we do not *encompass* an important area, we at least deal with it tangentially. Essentially, one cannot move *toward* a theory.

[2] An interesting analysis of this is to be found in Karl Mannheim's *Ideology and Utopia* (3). This is an extension of Marxian dialectic analysis applied to history and is in the area of sociology of knowledge. Other interpretations are found in the work of Robert K. Merton, Talcott Parsons, Max Scheler, and Pitirim Sorokin.

[3] Texts by popular writers of psychiatric interpretations such as Karen Horney, Robert Lindner, Theodor Reik, and others tend to do this very thing.

[4] In this connection, probably the most important attempt has been the development of the General Theory of Equilibrium. This theory states that for a given system composed of two or more elements, the average performance of the elements may be assessed. Then, it will be found that the performance of the *individual* elements may be specified as a direct *function* of their distance from the

THE THEORY OF *Deumbilification:* AN INTEGRATION TOWARD A NON-SPECIAL THEORY

Today, whenever one speaks of instincts, the academicians raise their noses. There is dissatisfaction with the concept, at least when applied to the human level, and this is justified in experience. An instinct is usually defined as complex unlearned behavior which arises without manifest practice, and it is evident that the study of man has shown time and time again that behaviors which were considered instinctive were not to be found in certain groups, or could easily be modified or prevented from arising. Thus, another illusion or explanation was lost; it was of no purpose to name a mystical force in man, and then explain by it. The mystical force just did not exist. However, *reflexes,* very simple automatic responses, are found. These mechanisms are studied by psychologists and physiologists, and the concept of the reflex has withstood scrutiny. Similarly, the *drive,* the generalized and diffuse activity in the given direction, has withstood the scrutiny of academicians. The more complex and specific term, "motive," however, is often suspect, and is receiving much attention in psychology at the present time.

If we have two acceptable concepts, reflexes and drives, what can be done with them? If behavior at the complex level occurs and is not instinctive, what explanations are satisfactory? Obviously, there must be some process of organization in the organism, not only physiological but also at the manifest level which we call mental. In this field, two other concepts, *maturation* and *learning-training* are considered legitimate in terms of the logico-empirical theory. Thus, we have a concept of development for man which begins in the fertilization of the ovum; the fertile ovum is fed and nourished in the womb. developing and becoming differentiated as an organism, prepared for some stimuli with some reflexes, and prepared to alert his older fellows to his drives by these reflexes. As we conceive it, ***the embryo is ready to learn as soon as it has established any behavior pattern.*** The one underlying mechanism of training, *conditioning,* is phylogenetically validated. Thus, we may conceive of the embryo in the womb as *capable* of learning. *In utero,* the organism is not subject to all types of stimuli, but certain facts are well established. The infant may be felt moving by the mother in the third month, so that it is well known that it begins exercising early in life. But aside from this, from pregnancy wastage, experimenters have found that the organism may respond to various forms of stimulation long before birth.[5] The one source of stimulation to which the organism definitely responds early in the

average performance. Further, if the direction of difference is maintained, *the sum of the differences will total to zero.* In no case will it be negative. The beauty of this theory is that *it has fitted all sets of data* to which it has been applied, irrespective of the sizes involved and irrespective of the type of distribution involved. The shortcoming of the theory is that while it is excellent for the description in the immediate, that is, the structural description, it does not take into account the sequence of structural descriptions which are the ongoing structural functional reality.

Recently, a mathematical model has been proposed for this theory. If the set of elements are called x_i, that is, x_0, x_1, x_2, and so on, and there are n such elements, then the expression for the sum of these can be stated as: $\sum_{n} x_i$. The average of these elements can be *computed* by dividing through the entire expression by n, since there are n such items. The expression then becomes: $1/n\sum_{n} x_i$. (Several steps are skipped here to simplify the expression.) A forthcoming paper will deal with this mathematical model, its extensions including the computation of the deviation from the mean, and possible applications to social science.

[5] The literature in this area has been reviewed carefully by Carmichael (1).

womb life is tactile stimulation. We may infer, and it is the thesis that underlies this paper. that *the child learns, although the learning may be small and generalized, in the womb.*

Sigmund Freud probably has done most to explain the important facts concerning man, but even he, being a pioneer, could not drive his analysis to fruition. We do not claim to present a complete thesis here, and certainly not a panacea for curing the problems of man, but we do feel that we are definitely breaking through previous barriers and extending in a more scientific manner work started by Freud.

Any serious student will see that *all* that Freud wrote is not acceptable. What we like in Freud is his approach, his attempt to find those situations which are early and which might well have a great deal to do with the personality of people, and for that matter. of peoples. Thus, instinct aspects of Freudian theory we do not need to maintain here. Similarly, concepts which are in respects pedagogical, such as the superego, the ego, and the id, may be dismissed as superfluous in this context. However, in doing this have we dismissed Freud? We have not, in fact, for his significant contributions have to do with the development of patterns of response, and reaction to the absence of stimulation when the patterns of response are established.

Probably the most significant concept developed by Freud is that of *penis envy*. The concept is one concerning the development of an awareness of a lacking on the part of females as they learn, either through sight or indirection, that they do not have the external genitalia the males do. The awareness itself is an admission of inferiority, leading to frustration, and potentially, to the redirection of aggression which we ordinarily call conflict. This has been well developed by Freud himself, and emphasized by his students (Adler). But what of the male? Is there a parallel situation for the male? Except for the facetious proposals of a few feminists who have stated that the male develops an awareness of lacking in not being able to *bear* a child, no serious suggestion has been brought forth. Why, then, is the male so closely associated with conflict? It is at this point that we must develop a theory which is *nonspecial*. In this, credit is due to Alfred Adler, whose emphasis on aspects of inferiority and superiority indirectly led to the discovery.

Often it is possible to overlook the obvious. Sex, as a root of problems, has been noted in the literature throughout the ages. But it took Freud to bring the obvious to the attention of the serious student! The scholar should not be blamed entirely, however, for repression and suppression are now known to account for much of this. Let us look at our last sentence once again and note the words repression and suppression. Might it not be possible that some source area of inferiority feelings is so completely repressed that we ignore it, obvious as it is? This, in fact, appears to be the case.

Freud, Rank, and quite a few others have directed attention to the prenatal period. None noticed, however, that in the characteristic position, the knee to head position, the umbilical cord of the fetus ranges and rubs against the fetus. Proportionately it is a large object, soft; but omnipresent. It may be wrapped around the fetus, or it may be caressing his face as he rotates in the womb. In any event, it is with him for the duration of his stay in the womb, and the *fetus is in constant contact with the umbilical cord*. This constant contact builds up expectation, through conditioning, of further contact. When the fetus loses the umbilical cord, an awareness of its absence is manifest. So far as is known, after parturition, in all societies and peoples the umbilical cord is removed from the newborn, either by cutting, biting, or letting it atrophy, as it does, naturally. *It is normal course for the newborn, among all peoples, to be deumbilificated.* The absence of the umbilical cord, and the memory traces associated with it, are the underlying reasons for the insecurity manifest in man.

With this knowledge, then, many things become immediately obvious. The perpetual seeking for a better condition may quickly be associated directly with the feeling of this insecurity, the memory traces of the umbilical cord being particularly awakened under given sets of circumstances. Similarly, anxiety and insecurity may be seen as the sources underlying aggression, and the expectation of these which are built into the very nature of man in his ontogeny explains the invariant history of conflict and warfare. But why should the association be with men? Obviously the association is again one which is determined in the situational development in the physiological context. Because of the differences implicit in the existence of the external genitalia, the *additional* feeling of insecurity (or inferiority) results in the ordinary suppression of aggression, so that females tend to be aggressive only in the more devious and protected ways. The *penis envy*, thus, *serves as re-enforcement of the insecurity condition*. Similarly, the locomotive restriction of being gravid and the dependence implied during gestation predispose for the development of more devious outlets for aggression among the females. One of these forms, of course, is the creation of conflict among others, and by quite natural grouping, among men. Thus, the organization of society may be expected in most cases to dispose toward conflict among men rather than among women, and this is a fact well verified in anthropological research. Conversely, males, having the external genitalia, occasionally supersede the feeling of insecurity by emphasizing the surrogate. *The penis may serve in some cases as the umbilical cord surrogate.* The memory traces, however, are not removed, and the constant presence of aggression leading to conflict and warfare is evident. The surrogate, the penis, leads to exhibitionism (as previously noted by Freud), and it is for this reason that we have the strong association among peoples of cults of beauty (almost synonymous with manhood), dancing, singing, expression in the artistic forms, and even the deities, with the male. A concept such as castration complex develops naturally with conditions associated with the positively re-enforced behavior.

We will not press the universality of the observations made here. However, consider the tremendous repression of umbilical reference in our society alone. *There is no known profanity with an umbilical reference!* No other ordinarily repressed or controlled area is this fully repressed.

On the side of symbolism, all societies are familiar, though they refuse to recognize them as such, with umbilical symbols. In many cases umbilical symbols such as the snake, the coiled snake, macaroni, and many others, have been identified erroneously as phallic symbols. A prime example is the identification of the male figure on the cover of the telephone directory as the paragon of phallicism when most obviously it is the epitome of umbilicalism.

On the side of behavior implicating memory traces and direct expression we have examples too numerous for detailed presentation. Consider the child's play with the umbilicus, or the religious example of Buddha contemplating his navel. Of particular interest is the universality of pleasure associated with caressing, and in particular, nuzzling, which are forms simulating the caressing of the umbilical cord. Nuzzling, it should be noted, has been demonstrated to occur through the phyla, while such behavior as kissing, nose rubbing, etc. has been demonstrated to occur only in *some* human communities. Further, while sexual intercourse is universal, pleasure is not necessarily associated with it, as is seen from frigidity studies among people, and passivity studies among females in various species.

One point of this brief paper is to indicate how the development of the nonspecial theory has already eliminated an existing theory and replaced it by more accurate description. A major contribution of Otto Rank was his theory of *birth trauma*, which essentially stated that generalized insecurities might be associated with the leaving of the womb, and in the pain and filth and gore involved in the extrication. While plausible, the situation did not become clear until recently. First, the large number of Caesarian operations has bred a population largely spared the "birth trauma," but no essential difference has been reported between this group and normal births. Thus, the special theory is demonstrated to be erroneous, and evidence is produced which is consistent with the nonspecial theory. It is evident that *the loss of the umbilical cord is the trauma.*

Further Developments: Mammary Envy

One of the most immediate reactions to the presentation of the nonspecial theory, which deals with the relationship of deumbilification to response characteristics, was the proposal that in fact the underlying source of *order* in behavior is attributable to *mammary envy*. The initiating proponent writes:

> I cannot develop the whole theory [of mammary envy] in this note of thanks for your lecture, but I am sure that you can see the many implications which can flow from the recognition that male and female human beings are, at various times in their lives, and in varying degrees, always *with* and *without* breasts. The meaning of this fact for early infancy is obvious, but it has the greater value of adding the developmental dimension to Freud. He never adequately appreciated the mature years. The unsuccessful efforts which have been made to develop the anal-oral-genital sequence may now be replaced by more fundamental analysis. Your contribution is a real breakthrough for me although closure is yet dimly perceived. I can sense, however, the enormous integrative power of a faintly conceived. pentagonal, three-dimensional paradigm ranging clockwise, and *irreversible*, from umbilicus on through anal, oral, genital to mammary.

The immediate reaction to the proposal was acceptance. This, however, did not prove fruitful. After considerable *empirical discussion* and *empirical thinking*,[6] it was discovered that certain consistencies and inconsistencies needed to be specified and accounted for. While there are feeding differences, on which a considerable amount has been written, implications of feeding are commonly translated into love nurturance and other concepts designating affective ties. Rejection of this type of analysis became necessary.

It is at this point that the *cross-cultural approach* became most useful.[7] Whole cultures differ in their feeding habits, and for this reason the personality complexes of individuals may be examined as a function of the feeding differences. In this way we found the clue of the relation of mammary envy to the nonspecial theory to lie in the family system of certain less mechanized cultures. Before introducing the datum from which we derived the clue we shall develop the relationship of person to breast.

First, let us acknowledge that a non-breast-fed baby may have to learn cer-

[6] Empirical discussion and empirical thinking are forms of research. Although not usually presented in most methodology texts, these research approaches have great prominence in actual practice.

[7] This is another example of the fertility of interdisciplinary research.

tain of the modes of response in relation to mammary envy through secondary sources, through the intellect, or possibly, through the belated stimulation of physiologically (and phylogenetically replicating) facilitated patterns.[8]

Let us then consider the breast-fed baby. Breast feeding is temporally *after* the umbilical stage, and the onset is almost concomitant with deumbilification. The child is nestled comfortably with the mother, and it has access to the nipple and the thin milk. Nestling involves the rubbing of the face of the infant on the breast, and other tactile stimulation. The hands of the infant, usually closed, however, are not involved. *The infant itself has no cognition of breasts of its own.* These are basic facts.

Among maturing girls in our society, there is some interest in the having of a reasonably ample bosom. A good point of reference in recognizing the passing into young womanhood is the development of the breast. Having the ample bosom is desirable, although the absence of it is not necessarily disastrous. In societies where breast feeding of babies is the usual thing, having an ample bosom may be one visible evidence of being able to nurture a family. Even up to recent times there has been some pride and but little shame associated with publically nursing the young. Occasionally, the having of bosom has been considered as a negative value, but never in any serious sense. We have in our history, for example, the pencil silhouette and other styles which have tended to constrict or hide the bosom for the female. These perverse conditions, however, rarely persist.

Unlike the umbilicus, there is little repression associated with the breast. The breast has been exposed in painting and sculpturing throughout history, and there is much reference to the breast in erotica. In American society of today the bosom is the constant point of attention for the more or less cheap pornography which is disseminated through Hollywood, pulp magazines, and other sources.

Thus, we see that there is a considerable importance attached to the bosom. Attention is of two kinds: first, in terms of the recognized function of nurturance of the young, and secondly, as an object of beauty and desirability. Now, we see that there can be such a thing as mammary envy between females in the naive sense that some females have and some haven't the appendages, and those who haven't may desire. However, this is not important. The important focus is in the *possessing* of breasts rather than in having them. We can immediately look at our cross-cultural picture and get the entire information required for analysis of the role the breast plays. The clue to the importance of mammary envy came when it was noticed that in terms of possessing breasts, *there are differences and these differences basically underly the familial patterns which are so varied throughout the world.* Thus, in a *polyandrous* society the female (as a daughter in the family) may be considered undesirable. However, desirability of the bosom is still evident in the fact that two (or more) men, usually brothers, may share the female. It is interesting that other animals may be highly prized in a polyandrous society, while females (who may be in a category with animals) may not be so prized. *Monogamy* we consider as a more or less restrained approach, primarily associated with sophisticated, highly developed, structured society which must control its

[8] Walking, for example, is not described as instinctive. However, when the organism is matured sufficiently physiologically, learning to walk may be almost instantaneous to the first trial.

more obvious "motivational" sources by artificial means. This is not to state that monogamy is associated only with higher societies in terms of technology and population density, because this is not the case. The *polygynous* family organization is extensive and is indicative of certain conditions. Polygyny is found throughout the world, whereas polyandry is associated primarily with areas where there is considerable deprivation in terms of the available resources in the ecology. But what is interesting about polygyny is that it occurs in those areas which are not so much known for the technical development but rather for sensitive art and control of emotion. The control may be at times posed by the society in terms of a hierarchy, or the control may be one which is associated largely with the tradition which is familial, though extensive throughout the social community. By the standards of which we speak, most of Western civilization would not pass as artistically sensitive or controlled in emotions. We find, thus, that societies which are polygynous are not those with which we are most familiar.

What the analysis of family forms led to was that *where polygyny is found, so is found the prizing of stock animals, particularly cows.* We come to recognizing the relationship. Cows tend to be worshiped and prized in the societies which are polygynous, and it is interesting that almost invariably cows and wives or females are interchangeable. More interesting is that *in many places the cow is more valuable than the wife.* This leads immediately to the difference between wives and cows, which is a *binary quantitative variable.* The principle of the possession of mammary glands was first recognized in this context. Breasts are the prized objects, and looked at cross-culturally, it can be seen that, where the *range of collection* is relatively unrestricted, cows and wives are both collected for their mammae, and cows may be preferred.

What became evident in the terms of the nonspecial theory was that while mammary envy was not operating as an underlying source of motivation, it did operate as a reinforcement of the general insecurity feelings which existed. It was found that where there was control of the collection of mammae, there was associated also a greater amount of insecurity, and this, of course, can easily be verified. We have already noted that in the polyandrous society people live at a subsistence level, and this, itself, is associated with a great deal of insecurity. Persons who possess many mammary glands are ordinarily those who have greatest security in terms of other forms of possessions as well. Even in a society such as our own we find that possession of mammary glands is to some extent associated with being a secure person. Economically, it is only the wealthy man who can afford a mistress, or even an occasional replenishing of the allowable mammae by divorce and remarriage. In terms of adjustment security, we find that mental illness is associated with being an unmarried rather than being a married person. The fact that it is a matter of possession of mammary glands rather than the having of them is attested to in that the mentally ill persons are more likely to be females than males.

Before passing on to the next important reaction and further contribution to the nonspecial theory, let us just mention in passing that symbolism of mammary envy has been neglected to some degree, even though the emphasis on mammary, such as is noted in our own society, is quite prominent. Very few people, for example, have recognized the importance of mammary envy in their desire (projected) to conquer a mountain. Consider the attention that climbing to the peak of Everest has brought upon two recent explorers.

FURTHER DEVELOPMENTS: DIGITAL GRATIFICATION

Another form of reaction to the nonspecial theory pointed to developments which are frequently associated today with fallacies in observation logic. If Freudian symbolism gives everything that is done phallic connotation, this detracts from the explanatory value of the phallic symbol. We have already indicated that there may be other items that have been grossly ignored.

Probably the greatest contribution that has been made in the development of the nonspecial theory of umbilicalism is pointing to the fact that in the sequence of umbilical to mammary there is concomitant another type of development which is again located strongly in the response to the environmental situation. This is a factor of maturation development and learning-training. We have noted in presenting the material on mammary envy that the infant suckles but does not have the use of his hands. What is of considerable interest is that *the facility with the hands is the things that characterizes humans.* This has been recognized in terms of the thumb, or the enormous dexterity that the human has in comparison to other species. However, what is neglected is the fact that this is something that develops over a considerable period of time and that *dexterity is something that grows almost with intelligence.*

What is most emphatic is that once again the emphasis on phallicism has probably obscured a great discovery. Most of the things which are associated with the penis are probably equally well associated with the fingers; that is, masturbation as a prime example of genital gratification is something that is associated with the hand. The emphasis has been so strong on the genitalia that actually there has been complete neglect of the gratification which is received through the tactual stimulation in the hands, here called *digital gratification.*

That digital gratification is a reality is something that we state rather than argue. One of the first responses of the infant child is that when he can no longer *possess* the breast, its substitute will be the thumb, or the fingers, or the hand, and he will place this in the mouth. It is not that the child gets gratification from sucking the thumb but that the child gets gratification in having his thumb in his mouth. *It is through the thumb that the child feels as well as through the mouth.*

That gratification is received through the digits is seen in the myriad ways in which caressing is manifest. Caressing may be self-directed, or it may be caressing of others. The same person may indulge in both. The common ground is that the person *caresses,* and it is through the digits that the person gets the gratification. We will not develop here the *Lenny Complex,* which was first noted in connection with John Steinbeck's classic work, *Of Mice and Men.*

Not only has genital gratification been used to mask this real relationship of digital gratification, but even oral gratification has been used or misused in this way by Freudians and neo-Freudians. We have already pointed to the fact that the infant gets gratification from his thumb rather than from his mouth. It is the object of oral attention which is desired, and not the mouth which merely serves as host. But consider that such things as nail biting and cigaret and pipe smoking are considered as oral gratification, when in fact they all involve primary usage of the hand. Consider also how few mannerisms are associated with the use of the mouth as compared with the hands; rubbing, fingering, thumping, drumming, etc. Expressiveness, when not located in the language or the histrionics of voice change, is most assuredly associated with the hands. In some cultures hand gestures are a secondary form of language.

Just recently an associate put his finger on an important example in this area by bringing up the story of Peter and the Dike. Peter's action, usually interpreted as an example of great courage and devotion, is actually, in the light of this new theory, one of gross self-indulgence.

In closing this section, let us recapitulate the previous two sections. First of all, we have indicated that the source, from which certain "motives" and forms of action stem, appears to be associated with deumbilification. Second, we have noticed that security is associated with the degree of *possession* of breasts. Third, we have found that the major source of positive gratification is associated with the digits. We have essentially replaced the Freudian sequence of anal, oral, to genital with the more appropriate one of umbilical, mammary, to digital. We have not destroyed the Freudian concepts; rather, we have shown that they were properly isolated, but insufficiently so, and that they must operate within the nonspecial theory. The so-called anal, oral, and genital stages, thus, are seen to be but *special* foci.

Further Developments: Reference Person Theory

The most gratifying reaction to the nonspecial theory has been an entirely independent contribution, and this has stemmed from work on the reference point of response.[9] In prior work, it has been noted that after a given event has occurred, it is possible to go back and indicate the reference points to which persons were responding. That is, if groups are known to have different sets of values, and behavior is associated with the values, then if we know that a person is responding as though he were a member of a given group, we may infer that he will behave as though he were a member of a given group. Thus far, in the relevant analyses, most of the prediction has been backward in time. That is, the inference is that because a person has behaved in a given way, he has considered himself as a member of the particular group which may be expected to behave in the particular manner indicated, or *at least*, he has responded in the same way as a member of the particular group which may be expected to behave in the particular manner indicated. This does not mean that he has behaved as all members in the particular group behave, because the different members of the group may behave in quite different ways, and he may be behaving exactly like one of the deviant members who does not act at all like any of the other group members. (The deviant, of course, might himself be acting as though he were the member of another reference group.) This type of analysis is quite difficult in the forward prediction except in relatively simple cases. What is of particular value in the development this far is the association of a person with a point of reference, not only for membership, but also for judgment. The behavior of a person, thus, includes his judgments, and these are a part of his reference point, and concomitantly, are relevant to reference group theory.

In our work we have been led to the reduction of the group to the diad, the simplest and smallest unit of two or

[9] A considerable amount of work has already been done in this area (4).

more persons. However, our intensive treatment has even made it impossible to work with groups of this size, and forcibly, we have been reduced to groups of size *one.* However, we are still working with a diadic situation, but we deal with the relationship of two one-person groups.[10] To prevent confusion of the diad with the situation of one-person groups, we have introduced the concept of the *person-group,* and in this connection, *reference person theory.* Once this distinction was made and we began working with this limiting case, it became evident that if we identify the reference person of ego, we may be able to predict his behavior in advance. In this connection we found that there was already, although it had never been brought to light before, direct relationship between reference person theory and the development of *self-exposition* of the asceto-physician.[11] It is exactly this relationship that tied the reference person theory to the nonspecial theory. As it turned out, it was not possible for self-exposition to develop until there was a complete rejection of the current symbol interpretation. It was not until the cloak of importance of phallicism could be removed that the asceto-physician could acknowledge that *inadequacies for persons occur in all spheres.* In this, it is a crucial point that they have called themselves asceto-physicians instead of psychoanalysts or therapists.[12]

Conclusion

Since this is already an abstract of a monumental work, we do not recapitulate. We only note in closing that much research is currently focused on the kinds of theory we have presented.

REFERENCES

1. Carmichael, L. The onset and early development of behavior. In L. Carmichael (Ed.), *Manual of child psychology.* New York: Wiley, 1946. Pp. 43–166.
2. Fromm, E. *Escape from freedom.* New York: Farrar, Rinehart, 1941.
3. Mannheim, K. *Ideology and utopia.* New York: Harcourt, Brace, 1936.
4. Merton, R. K., & Kitt, Alice. Contributions to the theory of reference group behavior. In P. F. Lazarsfeld & R. K. Merton (Eds.), *Continuities in social research.* Glencoe, Ill.: Free Press, 1950. Pp. 40–105.

(Received May 24, 1954. Early publication)

[10] Small group research has received much attention recently. Although there is no work published, we have experimentation with the *no*-person group. Scores are randomly selected according to random procedure, and are given random meaning. Data collected in this way are leading to much insight in the study of unlikely events. In this research, to forestall possible criticism of generalizations derived from laboratory experiments, we are using a no-way mirror setup.

[11] The normal course of asceto-exposition is long, and parallels many of the patterns of psychoanalysis. The asceto-physician must in all cases be an MD to practice. He must diet, expose himself to all ailments and maladies known, and constantly reduce himself to points near death. Patients may revulse on first sight, but they then become full of pity (cf. transference). This condition prevails until there is a *hardening* on the part of the patient, and he becomes indifferent to the asceto-physician. At this point, the patient, realizing how well off he is by comparison (reference person theory), and being accustomed to the deformed, demented, and degenerate asceto-physician, may face the world on his own.

[12] Asceto-physicians have refused the current name identifications pointing to the implicit acceptance of Freudian symbols in them. For example, psychoanalyst derives in three parts: psycho- anal- yst, one who has to do with an imaginary anus. Therapist is simply a contraction of: the rapist.

Vol. 61, No. 6 November, 1954

THE PSYCHOLOGICAL REVIEW

TRAUMATIC AVOIDANCE LEARNING: THE PRINCIPLES OF ANXIETY CONSERVATION AND PARTIAL IRREVERSIBILITY

RICHARD L. SOLOMON AND LYMAN C. WYNNE [1]

Harvard University [2]

The purpose of this paper is to describe two particular ideas, which we shall call "anxiety conservation" and "partial irreversibility," within a general theory of anxiety and avoidance learning. In doing so, more familiar postulates appear to us to generate some new and interesting theorems about behavior, some of which seem to correspond to established facts. We shall not concern ourselves at this time with an exhaustive review of empirical data. Rather, we shall direct our attention to a theoretical argument and shall illustrate specific points with observations drawn from the following fields: avoidance learning, psychotherapy, physiological psychology, and psychosomatic medicine.

Although we shall make no attempt here to validate thoroughly the logical deductions from our theoretical notions, we wish to point out that these conclusions are strongly suggested by empirical data reported in a series of our papers (39, 78, 79, 80, 87) as well as in other recent research reports (13, 14, 43, 58, 59, 75).

Pain-fear. It is a behavioral axiom that there are certain classes of stimuli capable of eliciting massive *pain-fear* reactions (see Miller, 59). These classes of stimuli are usually called unconditioned fear stimuli (see Mowrer, 62). The capacity of such unconditioned stimuli to produce pain-fear reactions is often presumed to be innately given (59), but some doubt about this has been raised by Hebb (32) in the case of a limited number of unconditioned stimuli. Whether innately given or not, the essential components of a massive pain-fear reaction may be characterized for analytic purposes as follows: (*a*) Autonomic nervous system discharge, resulting in visceral responses of high magnitude which are followed by feedback stimulation arising in the viscera and affecting the central nervous system; (*b*) skeletal motor discharge, resulting in diffuse and vigorous skeletal reactions which are followed by proprioceptive feedback stimulation arising in the musculature and joints and affecting the central nervous system; such reactions include aversive movements or

[1] We wish to thank our colleagues for many instructive and constructive criticisms of the preliminary drafts of this manuscript. We are especially indebted to G. W. Allport, Stanley Cobb, D. O. Hebb, W. O. Jenkins, G. L. Lindzey, O. H. Mowrer, and Eliot Stellar for valuable suggestions and criticisms.

[2] Much of the speculation in this paper leans heavily on impressions gathered from a project on anxiety in the Laboratory of Social Relations, Harvard University. This project is supported through the generosity of the Medical Sciences Division of the Rockefeller Foundation.

escape responses; (*c*) neuroendocrine discharge, resulting in the secretion of hormones which is followed by chemical feedback to the central nervous system and other physiological systems; and (*d*) higher central nervous system activity, as a direct consequence of afferent and efferent activity.

These four classes of phenomena will be assumed to be present whenever pain-fear reactions of high intensity are observed. Since our discussion will be limited solely to behavior which arises from very intense pain-fear stimulation (trauma), the reader may infer the presence of the four characteristics above whenever the words *fear* or *anxiety* are used in the following discussion. The parametric studies needed for the precise definition of intense trauma or intense fear are at present lacking. The absence of such a definition is a major weakness of this essay. However, we are temporarily willing to trust the intuitive judgment of the psychologist. We hope that most readers will have private conceptions of the attributes of intense fear as contrasted with weak fear, and we further hope that there will soon be some agreement in defining such a distinction.

Two conditioning processes. We shall maintain, as have several other writers (see Mowrer, 63; Schlosberg, 70; and Skinner, 77), that the facts of classical conditioning reflect a process which is not the same as the process of instrumental learning. Each process has distinctive characteristics (see Hilgard and Marquis, 33).

As applied to the learning of instrumental avoidance responses in the presence of intense pain-fear, the analysis of the two processes, we believe, should be made somewhat along the general lines suggested by Mowrer (62) in a provocative paper. We shall assume that pain-fear reactions become conditioned to previously neutral stimuli by virtue of a process of *Pavlovian*, or *classical* conditioning. The essential relationships of this type of conditioning as applied to fear reactions would be as follows:

> Any previously neutral stimuli which are followed closely in time by the occurrence of an unconditioned pain-fear stimulus, together with its immediate, elicited fear reaction, will eventually become capable of eliciting a fear reaction without the presentation of the unconditioned stimulus. The latter fear reaction is said to have become conditioned to the previously neutral stimulus (which is now the conditioned stimulus for a conditioned fear reaction). Following an earlier suggestion of Mowrer (61), we shall define the conditioned fear reaction as an *anxiety* reaction. The use of a term other than fear is justified here on at least two different counts. First, the conditioned reaction may have different components and different amplitude when compared to the unconditioned fear reaction; and second, the conditioned fear response is "anticipatory" in relation to the temporal sequence of events by which it is established. We assume the acquisition of an anxiety reaction to follow closely the empirical laws of Pavlovian conditioning.

The second process which occurs in the establishment of avoidance learning is that of *instrumental* conditioning. We may think of this type of conditioning as following either the laws of S-R reinforcement theory, of S-R contiguity theory, or of more recent cognitive learning theory. (For purposes of our exposition, we believe that it will make little difference which theoretical bias one might have.) The terms "anxiety reduction" or "fear reduction" may be translated for the purposes of this paper to signify stimulation change or termination. Applied to the avoidance conditioning situation, the relationships of instrumental conditioning may be described as follows:

> If a skeletal response occurs in the presence of an intense pain-fear reaction, and the characteristics of the response are such that, immediately subsequent to its occurrence, the intensity of the fear reaction is decreased, then

the skeletal response will be reinforced (its probability of occurrence will be increased in the presence of the recurring fear reaction). Thus, any conditioned stimulus which has acquired the capacity to elicit an anxiety reaction can subsequently acquire the capacity to elicit some skeletal act. One can assume that the anxiety reaction produces feedback stimulation with drive stimulus properties (see Miller, 58, and Mowrer, 63). Therefore, if some skeletal act results in removal of the conditioned stimulus for an anxiety reaction, the skeletal act will be reinforced. Anxiety reduction can be thought of as the reinforcing event for the skeletal act. But, before there can be any anxiety to reduce, the initial fear reaction in the presence of some unconditioned stimulus must have been conditioned to some neutral stimulus. An act which removes such a stimulus from the environment will decrease a conditioned anxiety reaction, and thus will be reinforced as an instrumental avoidance response.

Two sets of experimental conditions. The two types of conditioning require different conditions for the establishment of strong responses. The conditions which should be met for the *establishment of intense anxiety,* governed by the laws of *classical* conditioning, are as follows: (*a*) The intensity of the unconditioned, fear-eliciting stimulus, and the initial pain-fear reaction to it, must be great; (*b*) there must be reasonable temporal contiguity between the occurrence of the conditioned stimulus (previously neutral) and the occurrence of the unconditioned stimulus; and (*c*) the CS-US sequence probably must be repeated several times.

It should be emphasized that the three conditions above are needed for the establishment of high-amplitude (intense) anxiety in the presence of some specific conditioned stimulus. If something less is desired, certain aspects of the conditions above may be eliminated or altered. However, in our present discussion, we are only interested in anxiety reactions of a very intense nature. We are not convinced that repetition of the CS-US sequence is absolutely necessary for the establishment of an intense anxiety reaction in the presence of some conditioned stimulus, provided that the intensity of the original pain-fear reaction to the unconditioned stimulus has been very great (severe trauma). The number of repetitions necessary may also be affected by the "perceptual vividness" of the CS. However, we do know that if the three conditions above are carefully met, we will be able to establish a strong anxiety reaction occurring in the presence of some previously neutral stimulus.

The conditions which must be met for the establishment of very strong instrumental avoidance responses are as follows:

1. A skeletal act can terminate both the conditioned stimulus (which is a "signal" of approaching trauma) and the unconditioned stimulus; and furthermore, this skeletal act is of such a nature that it can *terminate* the conditioned stimulus before the unconditioned stimulus is presented, and it can *prevent* the occurrence of the unconditioned stimulus in the regular CS-US sequence.

2. The skeletal act must be followed closely in time by either *pain-fear* reduction (when the organism is "escaping" the unconditioned stimulus) or *anxiety* reduction (when the organism is "escaping" from the conditioned stimulus and is "avoiding" the unconditioned stimulus).

There is much experimental evidence which does not run counter to the analysis we have given so far. The work of Masserman (56, 57), Mowrer (63), Miller (58), Maier (55), and Liddell (44) provides many examples of the establishment of conditioned anxiety reactions *accompanied by* the development of escape or avoidance responses. While the exact interdependence of the two types of phenomena during acquisition is not yet known, Solomon and Wynne (80) have shown that conditioned anxiety reactions are apt to appear prior to the emergence of successful avoidance reactions. This was also observed by Mowrer and Lamoreaux (65). Such seeming independence of the two phenomena often reflects itself in very sudden acquisition of avoidance learning under the impetus of severe trauma. For example, Solomon and Wynne (80, 87), Brush, Brush, and Solomon (15), and Kamin (39) have shown several instances of a sudden transition from escaping to avoiding rather than a gradual shortening of latencies to the CS. Kimble (40) has reported the same phenomenon in rats. An example in Fig. 1 from the data of Solomon and Wynne shows this clearly. This dog quickly learns to escape from the shock, demonstrates a plateau, then suddenly acquires a stable, short-latency avoidance response.

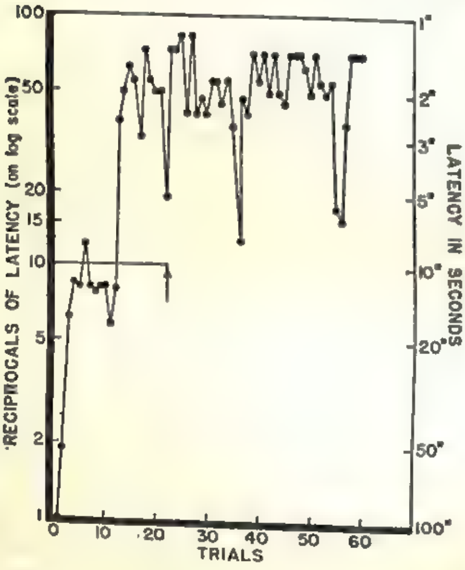

FIG. 1. A typical "sudden learner," from the data of Solomon and Wynne (80). The reciprocals of latency of response to the CS (in seconds) are plotted as a function of trials. The horizontal line marks the CS-US interval of ten seconds; points below this line are escape responses and points above it are avoidance responses. The arrow designates ten avoidances in a row. Note that this dog learned to escape quickly, achieved an escape plateau, then suddenly learned to avoid the shock by responding to the CS in less than ten seconds. Note the stability of the short-latency avoidance responses. They were even more stable after 200 extinction trials.

Avoidance learning is facilitated when the instrumental act not only prevents the occurrence of the US (avoids it) but also terminates the CS or danger signal (65). On the other hand, Bitterman, Reed, and Krauskopf (10) have shown that delay of termination of the US following an instrumental avoidance reaction (delay of reward) tends to strengthen rather than weaken the avoidance response. In their experiment they found that long shock duration is more important in intensifying the classical conditioning process than it is in weakening the growth of the instrumental response by virtue of delay of reinforcement. These authors consider their data to be strongly consonant with a dual-process theory of avoidance learning. Gantt and his collaborators have also noted the relative functional independence of anxiety reactions and instrumental acts (21, 26, 27).

We have no really strong convictions about the adequacy of S-R reinforcement theory, S-S contiguity theory, or S-R contiguity theory in handling the facts of the *development* of anxiety and avoidance responses. It will, we believe, occur to many that these theories may be applicable to many selected aspects of the data. But we do feel that the two-process approach to an analysis of anxiety and avoidance learning has helped us greatly in organizing for our own purposes a large body of experimental evidence. This, we feel, will become more apparent in handling the facts of *extinction* of traumatic avoidance learning.

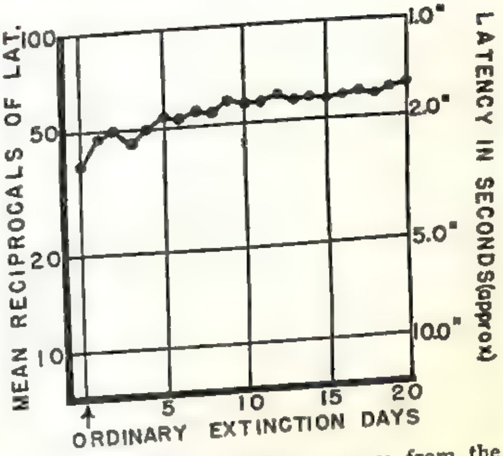

FIG. 2. A mean extinction curve from the data of Solomon, Kamin, and Wynne (78). Response latency as a function of unreinforced (no shock) trials. There were 10 extinction trials on each day.

AN ANALYSIS OF EXTINCTION OF INSTRUMENTAL AVOIDANCE RESPONSES

The strength of the learned instrumental avoidance response will presumably be related to the intensity of the classically conditioned anxiety reaction. Once the instrumental avoidance response is occurring regularly in the presence of the conditioned stimulus, then of course the unconditioned stimulus is omitted, and the temporal contiguity between the CS and the US is destroyed. When this occurs, the organism is terminating the conditioned stimulus instrumentally. It is important to note that once the organism is avoiding the unconditioned stimulus regularly, and so is not receiving any traumatic stimulation, we are meeting the conditions usually believed to be required for the *extinction* of a *classically* conditioned response. That is, the conditioned stimulus is no longer followed by the unconditioned stimulus, and so one would expect according to Pavlovian laws that the conditioned anxiety reaction would gradually extinguish. If extinction of the classically conditioned anxiety response actually occurred, then we would ordinarily expect the appropriate instrumental avoidance response to extinguish sooner or later. These events are more or less expected in terms of a two-process theory; but unfortunately, in traumatic learning, they often do not occur as prescribed.

Failure of extinction is difficult for any theory to handle. However, this phenomenon does occur, especially in traumatic avoidance learning. Sometimes the experimenter is impatient, and so sometimes the subjects in such experiments are merely characterized as having a *high resistance to extinction.* Whether one is willing to wait through the hundreds of trials often required for extinction of avoidance will determine one's views on the extinguishability of the instrumental acts. At any rate, it is clear that in traumatic learning we often do face some special problems in extinction. Mowrer (63) has reported extreme resistance to extinction of avoidance responses in rats; so have Miller (59) and Gantt (27). Masserman (56) and Maier (55) have made some related observations which corroborate Mowrer's impressions.

More recently, Solomon, Kamin, and Wynne (78) have made a rather molecular analysis of the behavior of dogs during extinction procedures following traumatic avoidance learning. They have shown that latencies (reaction times) of avoidance responses continue to shorten after the avoidance response occurs regularly. Thus, with steady omission of the US, the instrumental response of jumping a barrier became more and more stereotyped and the latency of response to the CS became more rapid (shorter), leveling off at 1.6 seconds. Dogs typically continued to respond to the CS for several hundred trials without signs of extinction. Figure 2 shows a mean extinction curve for 13 dogs. Some of these dogs had received only three or four shocks during a rapid acquisition sequence with a 10-second CS-US interval. The shock was extremely intense, just subtetanizing, during acquisition. Even with relatively long CS-US intervals, Brush, Brush, and Solomon (15) have shown that dogs in traumatic avoidance learning will "settle down" to short-latency responses to the CS in 200 trials after the last shock has been administered during acquisition. For example, with a 20-second CS-US interval during acquisition, the response latency asymptote approached 1.6 seconds, on the average, after 200 extinction trials.

Working with the writers, Kamin (39) has further shown that "spontaneous" jumping, in a free-responding avoidance learning situation, is most intense *after* the criterion of acquisition is reached and short latencies of response to the CS are observed. If frequent spontaneous instrumental acts can serve as a rough anxiety index, it is clear that cessation of presentation of the US is no guarantee of reduction of the strength of the conditioned anxiety reaction. Defecation, urination, and other ANS signs accompanied spontaneous jumping.

From our own data, and from those of others, it therefore seems clear that the extinction data of traumatic avoidance learning cannot be explained solely by existing principles. Existing theories would probably place emphasis on number and regularity of reinforcements, amount of anxiety reduction, lack of reality testing, strong expectations, stereotyped S-R contiguity, etc. But protracted extinction is not indigenous to any popular theoretical system. Two-process theory, as advocated by Mowrer (62), actually does contain some inkling of the resistance phenomenon, but it is not completely adequate in handling failure of extinction or certain observations on overt "emotionality." Gantt's (27) reasoning accepts the phenomenon of resistance to extinction but does not seriously try to explain it.

The expectations of the two-process theory of avoidance learning are portrayed in Fig. 3. Here we see that anxiety reduction reinforcement strengthens the instrumental avoidance response until anxiety is extinguished. Anxiety extinguishes because the CS is no longer followed by the US after the animal is successfully avoiding shock. Then, when the CS can no longer elicit anxiety, the instrumental act extinguishes because it is no longer followed by anxiety reduction. Thus, while Mowrer's notion does protract extinction somewhat, the phenomenon still should be readily obtainable. It may be that the fictitious state of affairs in Fig. 3 is a good model for *nontraumatic avoidance learning*. There are, however, some observations made by Solomon and Wynne

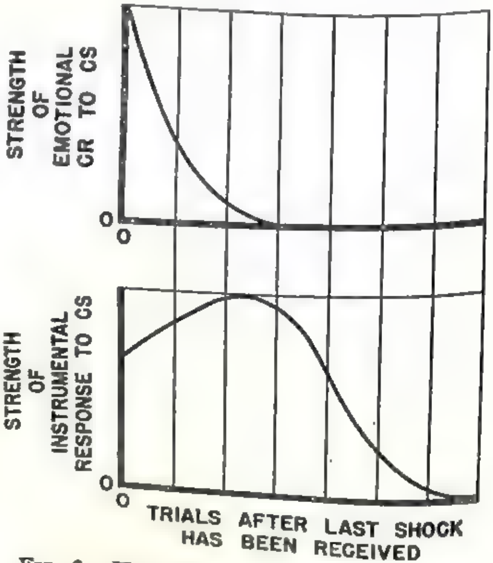

FIG. 3. Hypothetical relationships between intensity of a conditioned anxiety reaction and a learned instrumental response which terminates the CS for the anxiety reaction. These relationships apply to extinction trials only, and so the abscissa is in trials after the last shock reinforcement has been presented.

(80) and by Solomon, Kamin, and Wynne (78) which would render the simple two-process model inadequate for traumatic avoidance learning. These are the data to be explained:

a. We have noted the fact that, in return for a few intense shocks during acquisition of avoidance, dogs gave back as many as 650 avoidances without showing any signs of extinction. Others have observed essentially the same phenomenon (26, 44), and clinical evidence of a similar nature on human phobias exists in great abundance.

b. More important is the fact that overt signs of anxiety rapidly disappeared while the dogs were becoming more and more stereotyped in their jumping and their latencies to the CS were shortening. (If anxiety was being reduced by jumping, the anxiety reduction certainly was not evident at that stage.)

c. If anxiety was occurring covertly in the central nervous system, it did not obey any common laws of extinction, because if the dogs were forcibly prevented from jumping by means of a glass barrier they usually showed intense overt anxiety reactions (78).

d. If a dog happened to have an abnormally long latency on a particular trial, he typically acted "upset" immediately *after* the instrumental response had occurred, and jumped very quickly on the next few trials.

We do not think any popular theory could rigorously deduce these four phenomena; and our current favorite, two-process theory, will *not* do so. We feel that there are at least two important principles which might operate in producing all of these phenomena. We are not certain whether they are supplementary or independent principles. Let us say, then, that they are merely tentative ideas which will make two-process theory work better in handling the four facts of traumatic avoidance learning which we have selected and listed above.

The anxiety conservation phase. There is at least one very important possibility which has been completely ignored in the current analysis of extinction of avoidance responses. After the subject is responding to the CS with latencies shorter than the time required for the elicitation of the classically conditioned anxiety reaction, it is quite possible that at least the peripheral ANS part of this reaction will not occur at all. We hypothesize that the subject removes himself from the presence of the CS so rapidly that the CS is *almost* ineffective. The extent to which this is possible or likely for non-ANS components of anxiety reactions will be discussed later, in the section on physiological problems related to these ideas.

Figure 4 shows how intensity of the anxiety reaction might vary with time of presence of the CS, and how the peripheral anxiety reaction might subside after CS termination, as a function of time of presence of the CS. If the hypothetical events in Fig. 4 are approximately valid, then the occurrence of a rapid instrumental response to the CS would prevent peripheral anxiety reactions from occurring; *if nonreinforced exercise of a CS-CR relationship is the necessary condition for extinction,* then the extinction of the associational linkage between the CS and at least this portion of the anxiety reaction cannot take place. In one sense, the amplitude of the anxiety reaction is being *conserved* as a relatively intact potentiality, a latent functional entity. In common sense terminology, the subject is responding so quickly in the presence of the danger signal that he removes himself from its presence before he can become upset by it. Thus he never overtly experiences the set of events

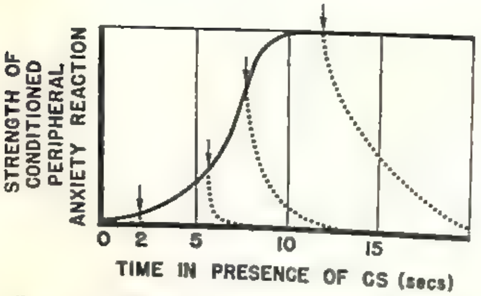

FIG. 4. Hypothetical relationships between time in the presence of an appropriate CS and the postulated strength of the ensuing conditioned emotional reaction. The arrows designate four points at which an instrumental act terminates the CS, and the dotted curves show the form of the anxiety decay curve for those four points. The term "strength" here is both quantitative and qualitative: it is assumed that more and more response elements of the ANS are recruited into the emotional reaction pattern as time in the presence of the CS increases. Thus, as time in the presence of the CS increases, the probability of long-latency visceral reactions being conditioned to the CS will increase.

necessary to "test reality." The anxiety reaction rarely occurs.

Under these conditions, however, the instrumental act *is not followed by anxiety reduction,* and so its habit strength will begin to decrease. When this decrease in habit strength of the instrumental avoidance response reveals itself in longer latencies, then the CS once again has enough time to elicit an anxiety reaction, and the instrumental avoidance response this time will be followed by anxiety reduction and a resultant increment in habit strength. Note how the long latencies in Fig. 1 are followed by rapidly shortening ones. In other words, the subject has "frightened himself" by not getting away from the danger signal fast enough, and he is "relieved" when he finally does remove himself from the signal. Simultaneously, however, there is weakening in the strength of the classically conditioned anxiety reaction, because the US has not occurred following the CS and the appearance of anxiety. The extent to which this weakening occurs will depend in part upon the intensity of the peripheral anxiety reaction, which, as shown in Fig. 4, is possibly a function of the time spent in the presence of the CS (the latency of the instrumental avoidance response).

It is interesting to note that the delay in the presence of the CS constitutes a *partial reality testing experience,* the effectiveness of which is probably proportional to the intensity of the emotional reaction that occurs during the delay. In a real sense we can say that drive strength, *D*, in Hull's (34) sense, is being "traded" for habit strength, $_{S}H_{R}$, as a consequence of such partial reality testing. During short, stable latencies of the instrumental avoidance response, habit strength is being sacrificed while drive (anxiety) is being conserved.

So far, then, we have accounted for facts *b*, *c*, and *d* above. But troublesome fact *a* still remains. In view of the foregoing theoretical considerations *extinction should occur* sometime, even though painfully resistant to ordinary extinction procedures. At some hypothetical moment, the CS should no longer elicit the anxiety reaction (owing to frequent partial reality testing) and then the extremely strong $_{S}H_{R}$ should begin to be overcome by $_{S}I_{R}$ (borrowing from Hull). The instrumental response is at this point no longer being reinforced by drive reduction. This is probably contrary to fact *a*.

Therefore, we do not really believe that ordinary extinction procedures must be effective in the case of severe trauma. Suppose, for a moment, we assume that traumatic avoidance learning, if terrifying enough, is completely resistant to *ordinary extinction procedures;* that, barring accidents in procedure, it is empirically possible to produce avoidance responses which will last for thousands of trials over a period of years. Our own observations (78) lead us to believe that this is, in fact, to be expected in dogs, though we are sheepishly aware of the fact that we haven't had the courage to stick with a dog for more than a few months of steady responding. Observations of Maier (55), Masserman (56), Liddell (44), Mowrer (63), and Gantt (26) bolster our feelings about this failure of extinction. So do recurring observations of persistent phobias in animals and man. Therefore, there must be a point at which the anxiety conservation phase is buttressed in some way; there must be some reason for such resistance to extinction as is represented in fact *a* above.

The principle of partial irreversibility of classical conditioning. In general, we feel comfortable with the analysis so far, *but with one extremely important qualification.* In the case of *intense* anxiety, established with the support of an initial, *intense* pain-fear reaction, we believe that the classically conditioned responses *have become incapable of complete extinction.* We are assuming that in such cases of intense anxiety (conditioned fear reactions), where the autonomic and skeletal reactions are of *great magnitude and involvement*, a *principle of partial irreversibility of classical conditioning* is operating. The principle we wish to describe is not that of total irreversibility, but rather that a "traumatic" or very intense "pain-fear" reaction taking place in the presence of some conditioned stimulus pattern will result in a *permanent* increase in the probability of occurrence of an anxiety reaction in the presence of that conditioned stimulus pattern (whenever it reoccurs).

This permanent change can be thought of as a decreased threshold phenomenon or as a sensitization phenomenon which is relatively permanent. A neurophysiological correlate might be the permanent reorganization of central nervous system networks. (These possibilities will be discussed later.) A physical analogy would be hysteresis. Such conceptions of partially irreversible changes (underlying behavioral phenomena) bear a strong relation to the concept of "the adaptation syndrome" which has been described by Selye and his colleagues (73). The adaptation syndrome is characterized by relatively permanent, partial reorganizations of hormonal functioning. We are merely generalizing the principle of partial reorganization from endocrinology to neurophysiology, and then to behavior.

Certainly this *general* notion is not original! Freud (23) uses the idea of partial irreversibility in describing the effects of trauma. (In fact, Freud also has an anxiety-preserving mechanism, repression, to go along with partial irreversibility. But perhaps this analogy is stretched.) Hull's (34) concept $_{S}H_{R}$ is established through the action of reinforcement (drive reduction), and it is preserved *as a fixed quantity.* Extinction is predicted by the fact that $_{S}I_{R}$ and I_{R} are subtracted from $_{S}H_{R}$ in the computation of $_{S}E_{R}$, effective reaction potential. White (86) and Allport (3) certainly would be astonished if a general concept of irreversibility were held to be novel. So, perhaps, might be Maier (55), Hebb (32), Kubie (41), and a host of others. But we would like to apply the partial irreversibility principle in a *very particular manner*, reserving it as a *property of the classical conditioning of "intense anxiety" reaction.* We feel that this specification of the general principle is in accord with evidence from a variety of sources and may be useful in generating psychological hypotheses which are more directly testable than has been possible with the *general* idea of partial irreversibility.

If this principle of partial irreversibility is taken seriously, it would mean that the *ordinary extinction procedures* of Pavlovian conditioning, when applied to a conditioned anxiety reaction of great intensity, will have *only a limited effectiveness in reducing the intensity of such a conditioned response.* The extinction procedures (characterized by dissociation of the CS and US) of classical conditioning should only be capable of diminishing the strength of a conditioned anxiety reaction down to some *irreducible minimum* in the presence of the CS, to some fixed threshold value which is above the zero point. The anxiety reaction will take place to some extent in the presence of a protracted conditioned stimulus; and even though extinction procedures are repeatedly employed over long periods of time, the anxiety reaction will merely decrease somewhat in intensity, never completely disappearing. Thus, the ordinary extinction procedures of *classical* conditioning will be only partially successful (see Fig. 5). The conditioned stimulus will always have the capacity to elicit an anxiety reaction of some magnitude. The irreducible, minimum, elicitation capacity of the conditioned stimulus will probably be a function of the intensity of the conditioned anxiety response before extinction procedures are started.

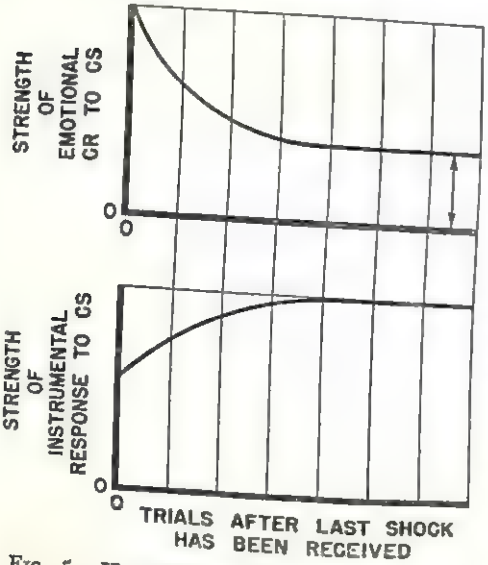

Fig. 5. Hypothetical relationships between the intensity of a conditioned anxiety reaction and a learned instrumental response for *traumatic* avoidance learning. Note that the asymptote for the emotional reaction extinction curve is above zero. The arrow designates the irreversible increment. Note that the strength of the instrumental avoidance response approaches some asymptote at a high value.

When we consider the significance of these assumptions for a theory of avoidance conditioning, a striking implication emerges. Without tampering with either the law of effect, the principle of anxiety reduction, or S-R contiguity principles, we can predict the *failure of extinction of instrumental avoidance* responses which have been established in the presence of intense pain-fear or anxiety. We arrive at such a conclusion, not on the basis of a drastic reformulation of instrumental learning and extinction laws, but rather, on the basis of partial irreversibility of *classically* conditioned anxiety reactions. There is no shortcoming of the law of effect implied here! If the latency of the instrumental avoidance response is long enough, the conditioned stimulus will elicit some degree of anxiety. Since the instrumental avoidance response has been established in the presence of anxiety, and since the instrumental avoidance response will continue to be followed by reduction of anxiety (by removal of the conditioned stimulus), the instrumental avoidance response will not be weakened. In fact, it will be strengthened through the action of the law of effect, approaching an asymptote of response strength long after extinction procedures have been instituted, long after the organism is successfully avoiding the unconditioned stimulus. Such a case is shown in Fig. 5. The principle of anxiety conservation leads us to expect that overt signs of anxiety will be expected to diminish, because the animal is not staying in the presence of the CS long enough to allow them to be elicited. This analysis now fits most of the troublesome facts which we have encountered in our own research (see facts *a*, *b*, *c*, and *d* above, as well as Fig. 2).

Retraining Techniques

Reality testing. Our theoretical fantasies also imply that forced reality testing extinction procedures may fail to eliminate the instrumental avoidance acts. Forced reality testing procedures are characterized by special manipulation, such that the organism is detained in the presence of the conditioned stimulus for long time intervals without the unconditioned stimulus being presented; and in addition, the organism is, by special means, prevented from exercising the learned *instrumental* avoidance response. It has been argued that this "forcing" procedure "lets the organism know" that the conditioned stimulus no longer "signalizes" the future occurrence of the unconditioned stimulus. Forced reality testing should result in "reorganization of the cognitive field" such that the organism no longer "perceives the situation as dangerous." Therefore, it has been argued, such a procedure should be effective in eliminating anxiety in the presence of the conditioned stimulus, and therefore, the instrumental act should subsequently weaken. However, application of the principle of partial irreversibility to forced reality testing leads us to predict that, while the anxiety reaction will be reduced in intensity, it will never be completely removed, and the avoidance response may therefore reappear even after long periods of reality testing.

We wish to emphasize that this theoretical position is *not* a doctrine of therapeutic hopelessness. We have pointed out that the originally intense anxiety reaction, elicitable by a conditioned stimulus, can be substantially reduced by the employment of traditional extinction procedures as well as the therapeutic procedures of reality testing. Some amelioration of anxiety is assumed to be possible, even in the case of most severe trauma. But a particular instrumental avoidance reaction, originally stemming from a severe traumatic event, will tend to persist if we merely employ ordinary extinction procedures or therapeutic reality testing.

Reward or "support." Therefore, it seems clear that if we are to eliminate a strong, learned *avoidance response* we must introduce some new stimulus conditions along with the other procedures above. But in doing so, we must keep in mind that the organism will, despite extinction procedures, always exhibit an anxiety reaction of some sort in the presence of the conditioned stimulus. One requirement would be that any added stimulation should have a high probability of eliciting a pattern of *skeletal responses*, which is *incompatible* with the occurrence of the avoidance response. Fatigue, competing avoidances, and competing appetitive reactions could be utilized. The stimulus pattern could be "motivational" in nature or could be composed of conditioned stimuli controlling strong responses which are incompatible with the avoidance reaction we wish to eliminate. Once the new stimulation conditions are introduced, and the resulting incompatible behavior is reinforced, the probability of the occurrence of the old avoidance response will decrease. The organism will then be capable of performing a new instrumental response in the presence of the conditioned stimulus for anxiety, yet the anxiety reaction *will continue to* *occur*. The old phobic or compulsive response will be gone, but not the anxiety reactions. At least, that is what we would expect if the principle of partial irreversibility of classical conditioning were valid.

Use of reward learning, or of appetitive motivation, in the elimination of strong avoidance responses has been studied to some extent, but the results are confusing. On the one hand, we have the excellent studies of Lichtenstein (43), in which traumatic stimulation was used to produce lasting feeding inhibition in dogs. And, on the other hand, we have those famous accounts of the curing of a child of fear of furry animals by bringing the feared object into the feeding situation (37, 38). Our guess is that the latter work represented a certain combination of forced reality testing, "crowding the threshold" of the anxiety reaction, *and* good luck! Unless the actual degree of anxiety is known, and the relation of its strength to that of hunger is known, it would be possible to produce neurotic inhibition of eating in the child by introducing the conditioned stimulus. This problem needs to be explored, especially with respect to avoidance responses that are established on the basis of intense trauma and overwhelming fear.

Our own observations are equivocal here. We have tried to retrain dogs in our traumatic learning situation by keeping them hungry five days and trying to use a feed lure to compete with the instrumental avoidance response. This was not a very successful procedure. Often the dog grabbed at the food and then performed the avoidance response. Much experimentation is needed in this area; especially needed are parametric studies which pit drives against each other in varying strengths, studies which might yield "isomotive curves."

Punishment. Another type of "new stimulation" which might be introduced in order to discourage the occurrence of an avoidance response is *punishment*. That is, when the organism responds to the conditioned stimulus with an avoidance response, some traumatic stimulus could be applied, preferably immediately following the occurrence of the avoidance response. Our theoretical ideas would deduce that such a procedure would only be effective *if* it is preceded by, or accompanied by, extensive forced reality testing. If punishment *for* avoiding is used before the classically conditioned anxiety reaction is partially extinguished, then the avoidance response might be substantially *strengthened* rather than weakened. Punishment would raise the general anxiety level, and since the avoidance reactions have consistently taken place in the presence of anxiety, it is conceivable that the organism would appear to be reacting more strongly to the conditioned stimulus than was the case before the introduction of punishment. However, it is conceivable that punishment could be introduced *after* forced reality testing, or after the anxiety reaction has been reduced to a low level, and that the action of punishment in this case might eventually lead to the extinction of the avoidance response. In this case, we would expect an initial strengthening of the avoidance response followed by gradual elimination of such instrumental responses. We are willing to believe that punishment which directly follows instrumental acts will tend to weaken the instrumental response strength (habit strength) of those acts. But in the case of punishment *without* forced reality testing, the avoidance response is being *negatively reinforced* on the one hand and *positively elicited* by anxiety drive on the other hand, and so it may appear to be unchanged or, possibly, strengthened by the action of punishment. In the case

TABLE 1

Mean Reciprocals of Response Latency for the Five Responses preceding and following the Introduction of Punishment for Avoidances

(From Solomon, Kamin, and Wynne [78])

Reciprocals of Latency	Punishment Introduced	
	After 200 Avoidances	After 20 Avoidances
Before punishment is introduced	50.7	50.5
After punishment is introduced	72.9	63.1

of punishment *with* forced reality testing, the anxiety level may be low enough so that the weakening effects of negative *reinforcement* (the Thorndikian stamping-out effect) can emerge after an initial period of raised anxiety level, resulting eventually in the extinction of the avoidance response.

Gwinn's experiment (30) demonstrates the expected phenomena quite clearly. In an escape learning experiment he found that, when shock was no longer given, a group of rats gradually ceased running. But an experimental group, shocked *for* running during extinction, at first ran *faster* for several trials. Some of these animals extinguished eventually, others did not. More recently, Solomon, Kamin, and Wynne (78), have shown the same phenomenon with dogs. In an avoidance conditioning extinction series, dogs were shocked *for* performing the instrumental avoidance response. They became extremely "upset," and performed the avoidance response with significantly *shorter* latencies. These data are shown in Table 1 for trials before and after the introduction of punishment for responding. In the same paper, it was reported that more satisfactory retraining occurred when punishment followed reality testing procedures than when punishment came before reality testing procedures.

Related Problems in the Psychology of Learning

Escape. One special case of learning initiated by noxious stimulation has some characteristics differing from those of avoidance learning. This learning type has usually been called escape learning or conditioning (33). It is characterized by simultaneous presentation of the conditioned and unconditioned stimuli, so that any instrumental acts on the part of the organism can never result in pain avoidance, merely in pain termination. Sheffield (74) has made a keen analysis of the events of escape learning contrasted with those of avoidance learning, and we are in essential agreement with his conclusions. He shows that in escape learning the presentation of the unconditioned stimulus tends to elicit unconditioned responses that are incompatible with the performance of a specific instrumental escape response which happens to terminate the CS and the US. For example, if a rat is required to run in order to terminate shock, and the shock level is fairly high, unconditioned crouching reactions may interfere with running. In addition, the rat may terminate shock only if he runs a given distance, so that the instrumental acts such as the initiation of running movements are actually punished for a finite period of time. Therefore, one would expect the development of a lot of diffuse responses (see Schlosberg, 70), some of which would actually be incompatible with a discrete and efficient instrumental escape act. When the unconditioned stimulus is later omitted in the extinction of the escape response, there would be present many response alternatives which might interfere with the previously reinforced instrumental act. Thus, according to Sheffield, the escape response would be weak, and might be easily extinguished when the unconditioned stimulus is omitted.

However, we would like to add that the general and diffuse reactions which have been elicited in the presence of anxiety would persist if the original stimulation were traumatic enough. We would expect the conditioned stimulus to continue to elicit some anxiety, just as was the case in avoidance learning, and some types of "aversive" skeletal reactions in the presence of this anxiety might be expected. Even though a very specific instrumental escape response may have been eliminated, the classically conditioned emotional reactions would tend to persist, though with an intensity considerably less than was the case before extinction procedures (omission of the US). Thus, the end picture of escape and avoidance extinction might be very much the same: partial persistence of anxiety, but accompanied by new skeletal reactions. At least, such would be anticipated by the principle of partial irreversibility of classical conditioning. We do not know of an experiment directly designed to demonstrate these phenomena, but such an experiment would be, in principle, a simple one.

Partial reinforcement. Sheffield and Temmer (75) have clearly demonstrated that escape learning, with shock level held constant, is less resistant to ordinary extinction procedures than is avoidance learning. They argue that the difference is due to the effects of partial or aperiodic reinforcement. Many animals show an irregular sequence of escape and avoidance trials when they are in the course of acquisition of instrumental avoidance responses. An animal may receive a shock on trial six because it did not respond with a latency less than the CS-US interval, and on the seventh trial it may avoid the shock, responding to the CS alone. Then, on the eighth trial, the animal may receive shock again for responding too slowly to the CS. Such alternation from shock to nonshock trials is common in many experiments. This is tantamount to aperiodic reinforcement if the US is considered to be the reinforcer of the instrumental act.

Now, based on a generalization decrement theory (36, 76) or a response-grouping theory (64), it would be expected that an aperiodic reinforcement schedule would lead to high resistance to ordinary extinction procedure when compared with the consequences of regular reinforcement. Escape learning is characterized by regular reinforcement, the US being administered on every acquisition trial. Regular extinction procedure merely omits the US and presents the CS alone. This is an abrupt transition from 100 per cent reinforcement to 0 per cent reinforcement, a condition which Sheffield and Temmer believe is conducive to easy extinction because of a generalization decrement—the animals can easily discriminate a change in the situation. But in avoidance training procedures the change from acquisition to extinction procedure is not as discriminable, and so the animals respond for more trials during extinction. (Another way of saying this is that the avoidance training procedure does not give the animals good cues for reality testing.) Sheffield and Temmer point out that early in ordinary extinction procedure the avoidance responses are actually less vigorous (of lower amplitude) than are the escape responses. These data contrasting escape and avoidance learning are paralleled by data in reward learning. In general, experimental work on partial or aperiodic reinforcement has shown that the irregular sequence of reinforcements leads to slower acquisition and slower extinction. With low ratios, extinction may well be a long and drawn-out procedure (36, 76).

In view of these considerations, it might be asked why we forsake a simple argument, such as the generalization decrement theory of extinction (76), in accounting for high resistance to extinction in traumatic avoidance learning. There are at least two reasons for our decision:

a. With an intense, traumatic US, the transition from escape responses to avoidance responses may often occur in a sudden fashion. An animal may absorb a few high-level shocks and then begin to avoid with perfection. This doesn't always mean that latency changes are abrupt, even though the escape-avoidance transition is. A good example of this is shown in Fig. 6, taken from the acquisition data of Solomon and Wynne (80). The same phenomenon was observed, even with very long CS-US intervals, by Brush, Brush, and Solomon (15). This is tantamount to a sudden shift from 100 per cent shock reinforcement to 0 per cent shock reinforcement, a condition which typifies the onset of ordinary extinction of *escape* responses. Such animals, learning to avoid very suddenly, do *not* appear to be less resistant to extinction than animals having a more aperiodic sequence of shocks and avoidances during learning. Thus, one cannot say that traumatic avoidance learning is typified by aperiodicity or partial reinforcement to the exclusion of perfect periodicity.

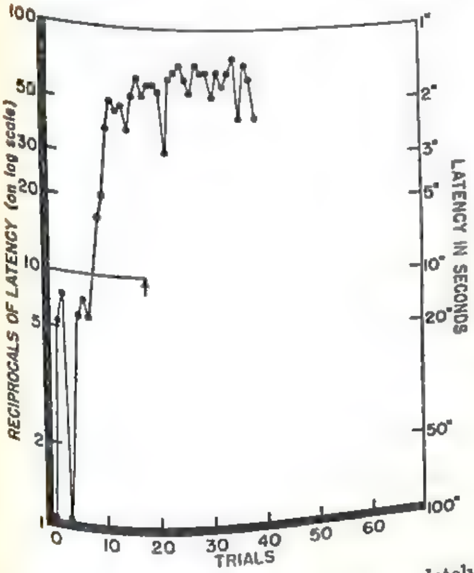

Fig. 6. A typical example of a completely sudden transition from escaping to avoiding, seen in traumatic avoidance learning in dogs (80). Note that while the transition from escape to avoidance is sudden, the latencies in this case change gradually after a sudden shift from a 15-second latency of response to the CS to a 6-second latency. This illustrates that the mere classification of instrumental acts into escape vs. avoidance, or percentage of successful responses, is not a totally adequate representation of the events taking place.

b. Escape learning is not comparable with avoidance learning because the CS-US intervals are not comparable. Escape learning is characterized by overlap of CS and US in time such that the US is unavoidable. But avoidance training procedure places a considerable period of time between CS and US so that the subject can react quickly enough to avoid the US. If the CS-US interval is an important variable in determining resistance to extinction, it would become a more complex task to assess the contribution of pattern of reinforcement as an important variable in traumatic avoidance learning. (This is a problem which we have just begun to explore, and the data are not conclusive at the present time. However, working with the authors, Kamin [39] has shown resistance to extinction to be a function of CS-US interval.)

In view of the data on resistance to special extinction procedures described by Solomon, Kamin, and Wynne (78), Masserman (56), Gantt (27), and others (44, 55), it seems reasonable that in traumatic learning the pattern or sequence of reinforcements is only one contributing variable of many in producing high resistance to extinction.

Nontraumatic learning. It has been suggested that perhaps all learning contains the element of partial irreversibility. Indeed, Allport's concept of functional autonomy (3) may be taken as a prototype for such a position. At present, we do not wish to extend our analysis beyond very intense, traumatic classical conditioning; but we realize that there is a possibility that other types of learning share some of the features of partial irreversibility of classical conditioning. "Law of effect learning," we feel, is totally reversible, and is even extinguishable "below zero" when the correct conditions are met.

In general, we feel that Gantt's (27) argument about the distinctive differences between conditioned autonomic reactions of an "appetitive" nature and conditioned autonomic reactions of the "emergency system" is cogent here. It can scarcely be argued that a Pavlovian conditioned salivary reaction—no matter how associated it might be with intense and eager orientation of the subject, and even overwhelming "expectations" of food (see Zener, 89)—involves the magnitude of visceral and hormonal reorganization that intense conditioned pain-fear reaction does. Nor is it too difficult to extinguish a conditioned salivary reaction below zero, as Pavlov (68) has shown; yet the conditioned heart-rate reaction to intense trauma has been shown by Gantt (27) to be incapable of total extinction in dogs, even over a period of years. While the conditioned appetitive reactions of reward learning may be a strong factor in influencing the resistance to extinction of rewarded instrumental acts, we do not feel that a partial irreversibility phenomenon need be involved.

There is, we feel (yet we cannot at present convincingly prove) a tremendous qualitative difference and quantitative difference between appetitive and traumatic learning. The exact nature of the differences remains to be explored. Perhaps it will point up the problem if we are reminded that a rat may starve to death rather than risk punishment, that a dog will refuse to eat if near a "danger signal" (43), that a sheep may exhibit hypertension due to trauma for months. (Perhaps Cannon [16] was making a distinction eventually useful to the psychology of learning when he delineated the *vegetative* and *emergency* functions of the autonomic nervous system. While this distinction now seems more dubious than Cannon made it out to be, since non-ANS functions operate, it seems premature to forget the distinction.)

RELATED PHYSIOLOGICAL PROBLEMS

During our discussions of the principles of anxiety conservation and of the partial irreversibility of intense anxiety reactions within the framework of learning theory, we have often found it illuminating to consider the compatibility of our ideas with present-day physiological facts and theory. We have felt that the consideration of the anxiety and avoidance learning problem from more than one starting point would suggest surplus meanings, and thus the possibility of additional modes of verification. Three main physiological problems are particularly relevant to this discussion. (*a*) What sort of physiological processes, among those aroused in intense anxiety reactions, are most capable of mediating skeletal avoidance responses? (*b*) To what extent do these physiological processes have discrete latencies, within which a prompt instrumental response could occur or begin? (These two questions pertain directly to the principle of anxiety conservation, which presumes that certain anxiety reactions could, on the one hand, motivate *delayed* avoidance responses, but, on the other hand, could be conserved from extinction by more *prompt* avoidance responses.) (*c*) To what extent do mediating physiological processes and associated anatomical structures have characteristics making them possibly susceptible to partially irreversible changes in the presence of traumatic stimulation?

As we indicated at the outset of this paper, we like to imagine that during intense anxiety reactions feedback stimulation occurs in at least four major physiological systems: autonomic nervous system, skeletal motor-proprioceptive system, neuro-endocrine system, and higher levels of the central nervous system. Since in the intact organism these systems clearly overlap and interact, and yet serve partly parallel functions, such a classification is highly arbitrary, but may serve expository and heuristic purposes. We shall consider, in turn, each of these four systems with respect to the three questions raised.

Autonomic nervous system and associated feedback. Discharge of the autonomic nervous system has long been regarded as prominent in anxiety. The existence of *efferent* pathways from various CNS levels to the peripheral ANS is now well-known and thoroughly established (24, 47, 84). However, if the autonomic aspects of anxiety are to influence avoidance behavior, anatomical and physiological mechanisms must exist whereby autonomic discharge arouses *afferent* feedback that actually can impinge upon projection areas and eventually influence skeletal motor centers and pathways. Only very recently has there been *direct*, detailed evidence that visceral *afferent* feedback may ascend in the CNS above the level of the lower brain stem (1, 4, 5, 9). It now appears that the neurophysiological characteristics of the visceral afferent projection system "differ in no important respects from the somatic projection other than in peripheral origin" (5, p. 457).

Neural feedback from autonomic reactions probably participates in both the specific and the diffuse projection systems to the thalamus and cerebral cortex. For the *specific* projection system, evidence is accumulating that there is an interaction of visceral and somatic afferent representation at several levels of the CNS (6, 24). Thus, the notion of the two-process learning theory that at least some of the afferent feedback impulses from the viscera have the properties of stimuli and so are capable of becoming conditioned stimuli and drive stimuli now seems reasonably consistent with physiological evidence. That is, not only may visceral reactions become elicitable by external stimuli through classical conditioning, but also the visceral feedback stimuli may become cues for instrumental acts, or may become motivating stimulation if their intensity is great enough (see Miller, 59).

Much recent research indicates that the *diffuse* projection system, involving the reticular activating system of the brain stem, is extremely important in emotional arousal, alertness, and the attentive processes (47, 54). The reticular formation probably also has a descending facilitating influence upon lower motor outflows (48). Although the relative significance of the various kinds of afferent stimulation that can activate the reticular system under different conditions is still not established, it is now very likely that feedback from both sympathetic and parasympathetic systems contributes to this afferent collateral network (22, 88). From the standpoint of avoidance learning theory, it is interesting that afferent impulses to the reticular system do *not* seem to function as cues or as specific conditioned stimuli. Rather, they seem to alter diffusely the alertness or attentiveness of the organism in perceiving and responding to impulses of the more specific projection systems (22, 81).

These physiological considerations, described, to be sure, in a highly schematic and oversimplified fashion, suggest the need for caution by learning theorists in discussing the properties of drive states and especially of acquired drives. Even the relatively well-understood peripheral autonomic nervous system, when it is active, results in afferent feedback impulses having a great potential range of variation in quality and magnitude, in duration and latency, as well as in the extent to which such feedback actually may influence instrumental motor responses. The loose term, "response-produced stimuli," while perhaps useful in highly generalized theory, tends to disguise these issues. And behind such issues may lie processes which greatly affect the extent to which anxiety and other learnable drives may actually mediate instrumental avoidance learning.

This discussion of autonomic feedback has emphasized neural pathways. However, it is clear that ANS discharge also produces widespread hormonal feedback affecting the reactivity of the organism in many diffuse ways. Some features of the possible role of such changes for anxiety and avoidance learning will be pointed out shortly.

To date, the various physiological and anatomical studies that have delineated the *details* of the mechanisms whereby autonomic reactions and feedback could influence instrumental responses have not been conducted under conditions of avoidance learning. Therefore, we do not know the extent to which the possible mechanisms are actually invoked in traumatic avoidance learning as discussed in this paper.

Inferential evidence concerning the *over-all* role of the peripheral ANS under actual traumatic avoidance learning conditions has been presented by Wynne and Solomon (87). They used surgical and pharmacological procedures in dogs to eliminate the sympathetic and parasympathetic peripheral nervous function. This was done in such a way that the capacity to perceive a traumatic US via *direct* sensory paths was unimpaired, and the motor capacity of the dogs to perform the instrumental avoidance response was not affected. In this experiment a CS-US interval of 10 seconds was used (allowing time for ANS reactions to develop in normal controls). Thirteen dogs who were given the surgical-drug procedures before training showed less uniform behavior than did normal control dogs. Ten of the 13 dogs were outside of the range of any normal controls, either in characteristics of acquisition or of extinction. In general, the experimental dogs were often retarded in reaching the avoidance learning criterion as well as in making their first avoidance response. However, they were all capable of achieving the criterion of avoidance learning and of responding with short latencies to the CS. Eight of these 13 dogs extinguished spontaneously. This had never occurred in any of 13 control dogs which had been trained under the same conditions. The experimental animals, as compared with normal controls, showed relative "indifference" to the shock, with few signs of autonomic and motor "upset," and relatively little stereotyping of responses during extinction.

Using tetraethylammonium with rats, Auld (8) has found that the acquisition rate for avoidance learning was depressed and that the extinction rate was accelerated. However, the pharmacological side effects of the drug make it difficult to interpret to what extent these learning changes were initiated by the effects of the drug on the peripheral ANS. A recent experiment by Brady (11) demonstrates these side effects to be extremely important in depressing general activity.

It was possible for Solomon and Wynne (80) to obtain only very incomplete data on ANS discharge and no objective data on associated feedback in their study of traumatic avoidance learning using normal dogs. Nevertheless, they observed a definite tendency for the first overt ANS reaction to the CS to occur approximately two trials before the first avoidance response. Mowrer and Lamoreaux (65) have reported similar observations. Fully satisfactory experimental evidence on the role of the ANS in contributing to anxiety as a mediating drive state in avoidance learning will not be available until direct measurements of both ANS discharge and associated feedback are possible under avoidance learning conditions varied along standardized parameters. However, the various kinds of inferential data now available quite clearly suggest that the peripheral ANS circuits have at least a highly significant, although not essential, role in avoidance learning under conditions of intense trauma during acquisition.

These various kinds of evidence suggest that, under certain conditions at least, feedback circuits which include the peripheral ANS *facilitate avoidance learning*.

According to the principle of anxiety conservation, the maintenance of avoidance under "ordinary" extinction conditions should depend, in part, upon the relation of the latency of the instrumental avoidance response to the total latency of classically conditioned anxiety reactions (efferent) plus feedback (afferent) which may be capable of facilitating the instrumental response. If we knew the latencies of the relevant physiological processes, we could predict anxiety conservation and prolonged maintenance of avoidance responses at a latency level just below, for most trials, the latency of these physiological processes.

In the case of the autonomic nervous system, the latency of measurable efferent reactions after the presentation of an external CS is ordinarily 1–4 seconds (2, 24, 47). It has been noted, for example, that in association with both light and sound stimuli a galvanic skin response of moderate electrical potential change has a latency of about 1.5 seconds (47). With the same mild stimuli, heart rate changes measured with the electrocardiogram also begin after about 1.5 seconds, reach a maximum after about 3 seconds, and have disappeared after about 4 seconds. Such measurable effector reactions presumably occur only after summation of less overt processes at both central and peripheral levels. These processes include, as examples, local intrinsic nervous reflexes in the gut in conjunction with local chemical and hormonal adjustments, mechanisms such as the carotid sinus reflex affecting cardiovascular and respiratory activity, and complex central circuits involving the primitive forebrain and the reticular activating system of the brain stem.

When the physiological processes brought into activity by external stimuli, especially stimuli arousing intense reactions, are considered in their full complexity, it is obvious that great caution is necessary in interpreting generalized statements about the latency of either efferent or afferent circuits. What can be said with some degree of assurance in the case of the ANS is that, in view of the demonstrable discrete latencies of peripheral ANS reactions, *effective* facilitation of skeletal responses by the feedback returning to central levels must also have a discrete latency. *Probably at least two seconds, perhaps several seconds, must elapse following a CS before feedback from the peripheral*

ANS can appreciably affect central motor processes.

These physiological latency characteristics of the ANS make the peripheral autonomic portions of anxiety reactions especially likely to be "conserved" by prompt avoidance responses which quickly remove the organism from the presence of the arousing CS. However, conditioned autonomic reactions do occur with a latency of only a few seconds and hence can be expected to help maintain avoidance when the instrumental response is delayed beyond this extent. The physiological properties of the ANS make it necessary to qualify such a generalization in several respects. For one thing, it is not correct to assume that the different parts of autonomic discharge and feedback all have the same latency. This point probably applies more to the latency with which autonomic reactions reach maximum intensity rather than to the latency with which they begin. For example, gastric reactions involving changes in acid production and motility probably take considerably longer to summate to maximum level than do heart-rate changes.

In other words, even within the autonomic group of anxiety reactions, a given duration of exposure to a CS may tend to elicit some kinds of reactions and to conserve others from full arousal. A recent experiment by Kamin (39) sheds a great deal of light on this problem. He studied the role of CS-US interval in a traumatic avoidance learning situation. The emotional reactions of his dogs varied with the length of the interval. Kamin's 5-second group showed a predominance of muscular tension, alertness, defecation, and urination during acquisition of a jumping response. His 20-second group, in contrast, showed a predominance of diffuse, agitated locomotion, retching, vomiting, and stomach rumblings which Kamin happily called "the gastronomic microphonic."

Such a contrast in symptoms must somehow be related to the "natural latencies" of visceral and skeletal emotional reactions. Some of the long-latency reactions probably are never conditioned when the CS-US interval is short. But of those reactions which are effectively conditioned, those with the longer latencies will be conserved during the earlier phases of extinction. It is clear that a detailed study of the qualitative changes in conditioned emotional reactions, during all phases of acquisition and extinction of avoidance learning, is sorely needed.

If partially irreversible changes do occur as a result of traumatic stimulation, such changes are most apt to occur in those structures in which the stimulation is convergent and most intense, or in structures with a particular vulnerability to such changes. Convergence might be especially likely to occur in pathways or areas with many afferent connections but with a relatively primitive or undifferentiated structure. In the case of the peripheral ANS, present anatomical knowledge does not suggest that such convergence is probable.

One might wonder whether the structural changes occurring in psychosomatic illnesses involving autonomic effector organs fall within the scope of the principle of partial irreversibility of intense classically conditioned anxiety reactions. There seems to be general clinical agreement that in individuals with such illness as duodenal ulcer, despite extensive changes in the person's life situation and despite prolonged psychotherapy, the ease of reactivation of the ulcer is greater than in the average individual. The extent to which such vulnerability can be due to traumatic life experiences aside from "constitutional" predisposition is highly controversial and largely speculative at present; "organ inferiority" is a disputed concept in psychosomatic medicine today. Also, it is conceivable that such psychosomatic diseases in autonomic organs might be secondary manifestations of partially irreversible changes in endocrine or CNS structures, rather than primarily in the peripheral ANS as such. This possibility is in accord with preliminary observations of Wynne and Solomon (87) which indicate that elimination of peripheral ANS functions *after* traumatic avoidance learning has already occurred has no effect upon the course of extinction. Dogs in this group were like normal controls in their extremely strong resistance to extinction of avoidance, in contrast to dogs in which the peripheral ANS was eliminated *before* acquisition of avoidance.

Endocrine reactions. Abundant clinical and experimental evidence indicates that endocrine reactions are capable of being classically conditioned. The relationship between emotional stimuli and the secretion of circulating epinephrine by the adrenal medulla has long been known. More recently, it has become apparent that the secretion of the antidiuretic, gonadotrophic, and adrenocorticotrophic hormones from the pituitary gland, as well as insulin from the pancreas, may also be influenced by conditioned emotional states (28). The adrenal medulla and posterior pituitary are controlled by a rich, direct secretomotor innervation (24), whereas the anterior pituitary (and thus indirectly the thyroid, adrenal cortex, and gonads) is probably under the neurohumoral control of the hypothalamus (31, 35). Various experiments indicate that all kinds of stimuli do not necessarily operate by the *same* mechanism to release the pituitary hormones, but that some, such as noise and restraint, require the normal attachment of the pituitary to the CNS, while others may act by way of circulating adrenaline or histamine, or even perhaps, via other metabolites or hormones (28). Thus, some specific understanding is gradually being built up of the various endocrine reactions which might be regarded as part of the pattern of anxiety reactions which accompany traumatic avoidance learning.

In general, the latencies of feedback from endocrine discharge, possibly except for epinephrine, are *considerably longer* than for the neural feedback from the peripheral ANS. Many of the *measurable* changes associated with adrenocortical discharge take one to four hours to reach maximum intensity (28), although the change may *begin* quite promptly. This means that a mediating effect of endocrine feedback within a given learning trial is likely *only if* the CS-US interval is much longer than that used in all reported avoidance learning experiments. However, such long intervals do occur in certain life situations of human beings faced with anxiety-evoking signals. Effective instrumental responses may not be possible for hours, days, or longer. In experimental avoidance learning an "acclimation" period may allow for endocrine reactions to occur to the total experimental situation. Also, with prolonged experimental sessions endocrine reactions aroused by early presentations of the CS may produce endocrine reactions affecting later responses. (May this contribute to so-called warm-up effects?) From the standpoint of avoidance learning theory, such considerations raise the question of whether endocrine feedback should be regarded as having "stimulus" qualities or as altering the reactivity of the organism to other, more specific stimuli.

Exactly how such endocrine discharge may "afferently" modify behavior and learning has not been studied in detail thus far. That a significant effect can occur is suggested by the frequency with which psychological aberrations are observed in patients with disorders of the endocrine glands (19, 83), and the frequency with which the administration of hormonal preparations results in deviant behavior (17, 69). However, the *specific* role of any hormone in avoidance learning processes is almost completely unknown. The difficulty in obtaining information on specific effects arises in great part from the impossibility of distinguishing where one hormonal regulatory mechanism stops and another starts. Regardless of the resultant behavioral effect, hormones act only by accelerating or retarding the rates of intracellular reactions which are catalyzed by specific enzymes (60).

There are only a few experimental studies of hormonal effects which are generally relevant to the anxiety and avoidance learning problem as we have posed it. Mirsky *et al.* (60), using monkeys and rats, have made observations in three different learning situations, including avoidance procedures, and have obtained results consistent with the interpretation that adrenocorticotrophic hormone (ACTH) *decreases* anxiety or its drive properties. Using sheep, Liddell *et al.* (45) found that administration of adrenal extract resulted in a rapid *reduction* in tension and a disappearance of the rigidities and tic-like behavior which characterized "neurotic" sheep. These authors did not use an avoidance training procedure. However, in seeming contradiction to these results of Liddell and of Mirsky, Applezweig (7) has found that hypophysectomy in rats interferes with avoidance learning even though the capacity for escape learning continues. When these rats were treated with ACTH, they showed partial restoration of avoidance learning.

About all that can be said about experimental work in this area in its present preliminary stage is that the pituitary-adrenocortical system seems to have some sort of effect upon avoidance learning, and presumably upon the secondary drive state mediating between CS and instrumental response; but the nature of such effects is obscure. It may be that the effects occur only under extreme experimental or pathological conditions, such as is the case with hypophysectomy.[3] Mirsky's investigations, inducing less extreme secondary physiological changes, deserve careful scrutiny because: (*a*) his results suggest that classically conditioned endocrine reactions may affect the course of avoidance learning by mechanisms which would be difficult to describe using current S-R or S-S formulations, and (*b*) because his results suggest that pituitary-adrenocortical reactions may counteract other kinds of classically conditioned reactions (negative, inhibitory, or inverse feedback?).

If endocrine reactions do, in fact, take part in classically conditioned anxiety, then the relatively long latencies of these reactions suggest that this portion of the total response may be strongly or inevitably subject to the principle of anxiety conservation by virtue of a prompt instrumental response. However, the tendency for such reactions to continue for long periods after the traumatic stimulus has been removed from the situation probably introduces a complication in the

[3] This may be unlikely. Recent studies of Gellhorn and his co-workers on insulin raise the possibility that hormonal mediating effects are powerful in maintaining conditioned responses. Gellhorn's most startling finding is that injections of insulin can result in the relatively permanent restoration of previously extinguished conditioned reactions (see Gellhorn, E. Is restoration of inhibited conditioned reactions by insulin coma specific for Pavlovian inhibitions? Contribution to the theory of shock treatment. *Arch. Neurol. Psychiat.*, 1946, 56, 216–221).

applicability of this principle to endocrine reactions to a greater extent than it does to ANS reactions. Nevertheless, short exposure to the CS could be expected to minimize the magnitude of the endocrine reaction and, in that sense, to conserve it.

The structural changes in many parts of the body apparently brought by intense discharge of the adrenal cortex in the "adaptation syndrome" (73) of Selye may, particularly in the "exhaustion phase," be closely related to the concept of partial irreversibility as we have formulated it here. However, even if such reactions *can* be classically conditioned and are partially irreversible, their pertinence to avoidance learning theory is *not* great *unless* it can be demonstrated that they have behavioral effects.

Furthermore, even if such behavioral effects can be shown under certain special conditions, it will still remain to delineate the range of learning conditions in which such considerations are relevant. A preliminary piece of evidence in this unexplored area is the finding of Mirsky *et al.* (60) that ACTH has no effect upon avoidance responses which have been thoroughly established. In contrast, the same avoidance response at an earlier phase of learning readily undergoes extinction with the same amount of ACTH. This suggests that under these learning conditions, at least, the extent of the effect depends upon the *phase* of the learning process. Possibly other structures may take over later in the extinction phase.

Skeletal motor-proprioceptive system. In discussing the physiological aspects of avoidance learning from the viewpoint of the two-process theory, it is necessary to differentiate artificially and arbitrarily the effects of proprioceptive feedback in facilitating continuous postural adjustments from the functions which such feedback may serve as conditioned stimuli or as acquired drive stimuli. Schoenfeld (71), in a recent review. emphasizes the role of proprioceptive stimulation in serving as negative reinforcement (punishment) or as *positive reinforcement* (reward). He says: "The proprioceptive stimuli produced by the avoidance response may. because they are correlated with the termination of noxious stimuli, become secondary positive reinforcers and hence strengthen the tendency to make the response which generates them" (71. p. 88). Schoenfeld feels that the concept of proprioception is superior to that of emotion and fear in describing the data of avoidance learning.

Certainly the physiological and anatomical properties of proprioceptive feedback are more clear than for any other physiological system. However, we are ignorant of the relationships between particular kinds of proprioceptive events and behaviorial events in avoidance learning. There are always millions of proprioceptive impulses impinging on the CNS at any time. The measurement of those impulses which are of particular importance for the development and maintenance of learned avoidance responses is, therefore, an overwhelming problem. This is not so much the case for ANS and endocrine functions in which particular alterations may be experimentally created and their effects on avoidance learning studied. It is certainly premature to pass judgment on the relative utility of concepts like fear and anxiety and proprioception in describing the data of traumatic avoidance learning.

Light and Gantt (46) and Loucks (51) have shown that movement of a limb and the consequent proprioceptive stimulation are *not* necessary for the development of a skeletal CR. Lauer (42) has further shown that a conditioned neural discharge to the dog's limb muscles can be established under total curarization. Thus, even when the dog was completely paralyzed, and distinctive proprioceptive correlates of muscular activity were therefore absent, avoidance responses were obtainable. These responses manifested themselves when the CS was presented after the effects of curarization had worn off. These observations do not rule out the possibility that proprioceptive feedback could facilitate, even though it is *not* essential to, the learning of avoidance.

It is well known that intense skeletal motor activity may lead to activation of the peripheral ANS and of many endocrine functions, with effects in the reverse direction also probable. Using dogs in which peripheral ANS function was eliminated before traumatic avoidance learning was begun, Wynne and Solomon (87) noted that these animals, compared to normal controls, showed much less motor "upset" and relatively little skeletal stereotyping of responses during extinction. This observation suggests that the ANS and its correlated feedback may have a more primary motivating role in such learning than does proprioceptive feedback. Perhaps the proprioceptive feedback may be inhibiting, contributing a work decrement factor.

Obviously, if the immediate motor reaction after presentation of a CS is an avoidance response, then other motor-proprioceptive processes which might be classified (more or less arbitrarily) as part of a total anxiety reaction, will not occur. In a sense, then, these other, later processes are conserved from extinction. This would apply to diffuse, classically conditioned skeletal responses and presumably not to other skeletal responses occurring at random. Schlosberg (70) has described the development of such skeletal responses.

Further analysis of some of these relationships may be possible using curare in dogs undergoing traumatic avoidance learning. An extremely important step in this direction has been taken in the ground-breaking experiment by Lauer (42), and we are currently following up this work by using curarized dogs in traumatic avoidance learning situations. The preliminary findings suggest that there is a rapid development of conditioned anxiety reactions in dogs under curare. However, the presence of such reactions does not necessarily result in rapid avoidance learning when the dog is later tested in normal condition. Rather, the skeletal effects of prior Pavlovian conditioning of anxiety under curare seem to be mostly composed of the diffuse postural adjustments and muscular tension emphasized by Schlosberg (70) in his account of classically conditioned skeletal reactions.

There seems to be no way of testing whether or not the motor-proprioceptive system may show partially irreversible changes in traumatic avoidance learning. If portions of this system in which partially irreversible structural changes might have developed are eliminated, the test response of instrumental avoidance is also eliminated. However, the rapidity of neural transmission plus the high degree of structural differentiation in this system make it seem a very unlikely locus of partial irreversibility.

Central nervous system. Within the CNS there are, clearly, numerous neural circuits which are aroused in massive pain-fear and anxiety reactions. Among these may be especially mentioned a well-defined circuit centering in the primitive forebrain which involves the hippocampus, fornix, mammillary body of the hypothalamus, anterior thalamic nuclei, limbic cortex, cingulum, and returns to the hippocampus (18). The primitive forebrain, loosely termed the "visceral brain" by MacLean (52), provides a system for integrating various intero- and exteroceptive impulses; complex autonomic functions are represented in a more organized manner here than in the circumscribed spot-to-spot representation of component autonomic functions in the premotor cortex (25, 53). Another important central network undoubtedly aroused in intense anxiety includes the connections between the reticular activating system of the lower brain stem, the thalamus, and the cerebral cortex (55). One or the other of these systems has been regarded by various authors (41, 47, 52, 67) as significant in the elaboration of "central emotion."

Direct physiological and anatomical evidence of connections with motor centers from primitive forebrain and reticular activating system is still incomplete. However, there do seem to be at least some association fibers from the primitive forebrain via the cingulate gyrus to cortical motor areas (85), and the reticular system apparently facilitates lower motor outflow (48). Major efferent discharge passes from the primitive forebrain via the hypothalamus to the peripheral ANS and to the pituitary-endocrine system (24, 53). Through such indirect mechanisms activity in these CNS association areas may affect skeletal motor responses in addition to the more direct CNS connections to motor areas.

Because of the difficulty of making direct measurements or observations of activity in such CNS circuits in response to a CS, these CNS reactions are awkward to describe as classically conditioned responses. Yet, if these CNS circuits can be regarded as the loci of "mediating" reactions which are activated by an external CS and which facilitate an instrumental CR, *then they are basically comparable to the peripheral acquired drives* described, for example, as occurring with ANS reactions.

Such an interpretation seems to be substantiated experimentally by the findings of Brady *et al.* (12). These authors have shown that cats receiving bilateral lesions of a portion of the primitive forebrain, namely, the amygdaloid complex and overlying cortex, acquired avoidance behavior of a given level with a significantly lower frequency than did operated or unoperated control animals. In this study the elimination of the ability to learn to avoid was incomplete and varied from animal to animal. This indicates that this part of the primitive forebrain, while not essential, probably is significant in the learning of avoidance. Whether a more complete elimination of the capacity to avoid would be obtained by ablation of other or greater portions of the primitive forebrain, or of other parts of the CNS, is still problematical.

Schreiner *et al.* (72), using the same cats as those in the avoidance learning study of Brady, observed marked changes in the sexual behavior of these animals which seem to be related to their endocrine functioning. These authors, therefore, suggest that the observed changes in behavior, precipitated by injury of the amygdaloid complex, may be due to a postsurgical state of altered endocrine activity. This possibility points up a difficulty in the interpretation of all ablation experiments in the study of avoidance learning, namely, that observed behavioral changes may be related to indirect and secondary changes produced by the ablation. However, the fact of the change suggests some sort of involvement of a specific brain area in the total sequence of physiological processes which underly the behavior being studied. A vast amount of work will be necessary, under standardized learning conditions, with various combinations of procedures which eliminate or alter various physiological functions, singly and in various combinations, before it will be possible to make full interpretations of the physiological bases of such learning problems.

Little pertinent data are available concerning typical reaction latencies within these CNS circuits which have been thought to mediate "central anxiety." Part of the total latency, in the case of the reticular activating system, has been measured as 0.4 second, for the blocking of the alpha rhythm of the EEG after the onset of a sensory stimulus (47). Thus, these CNS latencies are probably discrete, though most likely a good deal shorter than for some of the peripheral circuits. Theoretically, such relatively rapid latencies in the CNS might reduce the applicability of the anxiety conservation principle. However, a prompt avoidance response would still reduce the duration of exposure to the CS and so would reduce the proportion of anxiety which is elicited and subject to ordinary extinction.

In our formulation of the principle of partial irreversibility of intense classically conditioned reactions, we have spoken of the phenomenon in terms of decreased threshold or sensitization. We have hypothesized that such changes would be especially likely to occur in those structures which are subjected to intense, convergent stimulation but which have a relatively primitive structure for differentiating such stimulation. Various authors have speculated concerning the possible nature of such changes in the CNS. MacLean believes that "it is possible that if a certain electrical pattern of information were to reverberate for a prolonged period or at repeated intervals in a neuronal circuit, the nerve cells (perhaps, say, as the result of enzymatic catalysis at specific axon-dendritic junctions) would be permanently 'sensitized' to respond to this particular pattern at some future time" (52, p. 349). Hebb has discussed the possibility that structural changes in the form of synaptic knobs may develop with intense or prolonged neuronal excitation (32).

Although the reticular activating system is probably subjected to a great variety of intense afferent stimulation under traumatic learning conditions, the structure of this system is apparently highly complex and differentiated. Therefore, we would not expect that the reticular system would be a major locus of partially irreversible changes. Because ablation of this system results in loss of consciousness of the organism, the possibility does not seem amenable to test using this experimental procedure. More promising is the primitive forebrain. This area, as we have already mentioned, brings into association a great variety of sensations, especially visceral feedback, which can be expected to be prominent in intense anxiety reactions. At the same time, the poorly differentiated cortical cytoarchitecture of the primitive forebrain (49, 50) suggests that it probably has little efficiency as an analyzer or discriminator of these convergent visceral and emotional functions.

These speculations are of especial interest because they seem to be in accord with actual experimental data of a preliminary sort. Brady *et al.* (12) found that cats which were given bilateral amygdala lesions *after* acquisition of avoidance showed a slight postoperative decrement in avoidance behavior. However, for the three animals in this experimental group, the decrement was not statistically significant. In the same study, Brady found that three other cats with lesions in the orbital-frontal area, which is closely related to the primitive forebrain, showed a complete loss of previously well-established avoidance responses. After two days of postoperative retraining these animals again reached the previous avoidance response level.

We have felt that partially irreversible changes in traumatic avoidance learning are related to the magnitude or intensity of convergent stimulation. Elimination of an appreciable portion of afferent feedback during acquisition of avoidance should interfere with the *development* of CNS partial irreversibility, but not affect partially irreversible changes *previously* established. The results of Wynne and Solomon (87) using procedures which eliminated peripheral ANS functioning are in accord with these inferences. They found that when these procedures were used *before* training was begun, avoidance behavior was much more easily extinguished than in normal control animals. However, if these procedures of ANS deprivation were carried out *after* avoidance had been learned with an intact ANS, then *no* effect upon the avoidance behavior was discernible. To be sure, these experiments did not determine what structures or processes had taken over in the latter instance; certainly CNS loci are most likely. Combinations of ablation experiments could test this possibility.[4] Thus, the physiological correlates of such hypotheses concerning avoidance behavior come to have empirically testable consequences which are intimately related to a potentially coherent theory of avoidance learning.

PROBLEMS RELATED TO LIFE HISTORY STAGES AND PSYCHOSOMATIC MEDICINE

The phenomenon of partial irreversibility is assumed to be *characteristic of classical conditioning only;* it is thought to result from massive feedback as a consequence of reactions to intense trauma. In thinking about the implications of such a principle, we felt that there might be some characteristics of infantile conditioning from which one could logically deduce some of the phenomena of early learning which various psychoanalysts, and in a different framework, Hebb, have discussed.

Freud (23) has maintained that early trauma produces long-lasting influences on behavior. In his conception of "primary anxiety," early traumatic experiences, mainly birth and infantile deprivation, are the basis from which later anxiety develops. Drawing upon a variety of clinical and experimental evidence, Greenacre (29) and Stern (82) have modified and elaborated earlier psychoanalytic formulations of this problem. Two of the properties of primary anxiety are a wide range of reactivating situations and a pervading enduringness of the tendency to reactivation. In contrast, postinfantile, situational anxiety has as two of its properties relative restriction to specific situations, and relative transientness. Hebb, approaching the same problem from the point of view of perceptual learning, has pointed out the enduring character of early perceptual organization: "Phase sequences" may be immutable through time when strongly established (32). We believe that a two-process learning theory, combined with the principle of partial irreversibility of classical conditioning, will deduce these phenomena directly.

In the first place, the infant is instrumentally helpless when contrasted

[4] The reader will probably note the absence of a discussion of the effects of frontal lobe ablations on anxiety reactions and avoidance responses. While such a discussion might be appropriate here, we felt that this would get us into special neurological problems which would take us too far afield. For example, it still is not clear to what extent the effects of frontal lobe ablations are an index of frontal lobe tissue damage, damage to connections between the frontal lobe, thalamus, and limbic system, or retrograde damage in lower brain centers themselves.

to the adult. Thus, there is a greater susceptibility to *intense* stimulation and a probability that specific CS-US sequences will be repeated many times. The infant may not have instrumental acts in its repertory with which it can terminate unconditioned stimuli quickly, or escape from conditioned stimuli. Thus, classical conditioning should be greatly facilitated in infantile stages. Pain-fear and anxiety reactions should be common, then, accompanied by diffuse skeletal reactions which are not particularly successful. This is not to state that the infant cannot control its environment to some extent. Rather, we think that the infant cannot do it as well as, say, the adolescent; and this leaves him, to a *relatively* great extent, a pawn of certain environmental stimulus sequences. In addition, early conditioning would be characterized by broader stimulus generalization, due mainly to the lack of conditioned discriminatory reactions and precise verbal symbols. Then, too, it is possible that a "new," possibly more "plastic" nervous system might be much more likely to show large-scale reorganization in the face of traumatic conditions. All of these considerations lead us to believe that early traumata are more likely to produce partial irreversibility of classical conditioning than would later traumata.

Later traumata will have somewhat different consequences. Here we can expect quick instrumental action, so that the repetition of CS-US pairings would be greatly cut down. We would expect a lot of conditioned avoidance reactions of a fairly discrete nature to develop, but no extremely strong classically conditioned anxiety reactions *in most cases*. With overwhelming trauma, however, we might find partial irreversibility with a single CS-US pairing, accompanied by avoidance responses of a specific sort. Clinical evidence suggests that this may happen with catastrophic war experiences, for example.

From one point of view, what we have suggested is that early, severe traumata are likely to produce classically conditioned emotional responses of a lasting sort. Later severe traumata are more likely to produce a variety of instrumental acts of a high strength. Visceral and hormonal conditioning should, therefore, be more characteristic of early traumata and be elicited by broadly generalized situational stimuli. Neurotic symptoms, such as phobic acts, compulsive behavior, etc., should be characteristic of later traumata, and they should have high specific symbolic content. If we accept such speculation, we then are led to conclude that many psychosomatic disturbances would be derivatives of early traumata, whereas the predominance of neurotic symptoms (avoidances, phobias, compulsions) should characterize later traumata.

One additional characteristic of later trauma should be considered. Verbal or symbolic capacity would greatly complicate stimulus generalization which takes place with reference to conditioned stimuli. Depending upon the attitudes of the individual, the prior categorizations of events, and degree of differentiations available at the verbal level, the meaning of a CS-US sequence could be quite complex. Reality testing would also be complex, therefore, and might be carried on symbolically to a great extent. This complex differentiation would be less the case in instances of infantile traumata. Stimulus generalization would follow inherent gradients, would be relatively free of specific symbolic interpretations. Reality testing would be more closely bound to extinction principles, and would consist mainly of dissociating conditioned stimuli from unconditioned stimuli. Thus, we would expect reality testing in later traumata to be highly verbal; reality testing for early traumata would be highly inarticulate.

Related Psychotherapeutic Problems

It is interesting to note some of the implications of the principles of anxiety conservation and partial irreversibility for the appropriateness of psychotherapeutic procedures for various kinds of disorders. Psychotherapy relying largely upon the environmental manipulation and alteration of obsessive, compulsive, or phobic behavior would, using our formulation, seem to depend upon the substitution of one instrumental response for another. Such therapeutic procedures might bring about more socially desirable behavior, but could not be expected to remove completely the classically conditioned anxiety reactions which elicit one form or other of the neurotic behavior.

Another therapeutic approach might involve the introduction of competing motivation or the introduction of stimulus situations which elicit strong responses that are incompatible with neurotic avoidance responses. But the individual undergoing such therapeutic procedures would still "feel uncomfortable or anxious" in certain situations. This would be true even if the whole cognitive organization of the individual had been drastically modified. In a sense, all the various cognitive and manipulative procedures can be regarded as conserving anxiety from extinction processes. For example, compulsive hand washing is typically carried out before anxiety over real or fantasied dirt is fully experienced. When such patients are prevented from washing their hands under such stimulus conditions, marked anxiety or panic may ensue. From this standpoint, the skilled psychotherapist is able to help the patient to let his anxiety be re-experienced in a supportive setting (so that the therapist is not thereafter avoided in common with other panic-provoking stimuli!).

Such a therapeutic procedure is in accord with the principle that classically conditioned reactions must actually occur if they are to undergo extinction. A considerable decrease in the intensity of symptoms can be expected if the therapist establishes a relationship with the patient which duplicates parts of the conditioned stimulus pattern associated with original traumatic experiences. Optimally, an intense transference relationship can make possible nonverbal reality testing and the *partial* alleviation of anxiety having preverbal origins.

However, according to the principle of the partial irreversibility of traumatic anxiety reactions, there will be certain definite limitations on the "curing" of behavior arising from early, "primitive" traumatic experiences. This will also hold true for psychosomatic symptoms which may be a more direct manifestation of early conditioning. Complete freedom from a *tendency* to manifest such symptoms could not be expected, even with the most advantageous course of therapy.

(The authors have been informed by several individuals who have personally undergone prolonged and successful psychoanalysis that conscious efforts to relive in fantasy their most anxiety-provoking life experiences still reactivate a discernible residual of this anxiety which was once much more considerable in degree. However, these individuals feel they have sufficiently learned to discriminate these anxiety-provoking situations of early life from those situations ordinarily occurring in adulthood. Such distinctions, of course, involve gradients, not sharp demarcations.)

SUMMARY

We have presented a highly speculative analysis of anxiety and avoidance learning. The central assumptions of the analysis were: (*a*) there are two basic acquisition processes, the classical conditioning of emotional reactions and the instrumental learning of skeletal responses; (*b*) the laws of emotional conditioning are those of Pavlovian conditioning, where the relationship between a CS and a US is of prime importance; (*c*) the laws of skeletal response modification are those of reward and punishment learning, a Thorndikian or Hullian conception; (*d*) in the case of avoidance conditioning, the analysis of extinction phenomena can be more easily made if we introduce the principle of "anxiety conservation," a theoretical label for complex interrelations between the strength of anxiety reactions and instrumental avoidance responses; the central idea in the principle of anxiety conservation is that short-latency avoidance responses will prevent the CS from arousing anxiety reactions, thereby conserving conditioned anxiety reactions from extinction; (*e*) in the case of intense trauma as the US, the classical conditioning of emotional reactions is partially irreversible, and extinction of such reactions can only be partial at best; the central idea in the "principle of partial irreversibility" of emotional conditioning is that feedback from peripheral ANS reactions results in an overloading of rather primitively differentiated areas of the "emotional brain."

The consequences of such an analysis have been discussed with an eye to both behavioral and physiological research on anxiety and avoidance learning. We have included highly selected applications of our analysis to research on avoidance learning, psychotherapy, and psychosomatic medicine. We have specifically worked out implications of the analysis for: (*a*) extinction of anxiety reactions, (*b*) extinction of instrumental avoidance responses, (*c*) the effectiveness of reality testing therapeutic procedures, (*d*) the use of reward and punishment in facilitating the extinction of instrumental avoidance responses, (*e*) the effects of trauma at various stages in life history, and (*f*) the limitations and appropriateness of certain psychotherapeutic methods in alleviating anxiety and overt neurotic behavior sequences. We have not hesitated to introduce very speculative physiological considerations wherever we felt that it might stimulate new research or suggest new approaches. The role of the peripheral ANS, the endocrine system, the proprioceptive system, and certain areas of the CNS were discussed relative to anxiety reactions and avoidance learning.

In conclusion, the authors feel that they have not enunciated a "theory." Rather, they have presented some ideas and speculations which have been personally useful in organizing some facts of psychology. Some of these ideas have empirical consequences and so the wildness of speculation may perhaps be excused.

REFERENCES

1. Aidar, O., Geohegan, W. A., & Ungewitter, L. H. Splanchnic afferent pathways in the central nervous system. *J. Neurophysiol.*, 1952, **15**, 131–138.
2. Allen, W. F. Studies on irradiated cerebral differentiated excitation and inhibition as indicated and measured by respiration. *Amer. J. Physiol.*, 1942, **136**, 783–795.
3. Allport, G. W. *Personality: a psychological interpretation.* New York: Holt, 1937.
4. Amassian, V. E. Cortical representation of visceral afferents. *J. Neurophysiol.*, 1951, **14**, 433–444.
5. Amassian, V. E. Fiber groups and spiral pathways of cortically represented visceral afferents. *J. Neurophysiol.*, 1951, **14**, 445–460.

6. Amassian, V. E. Interaction in the somatovisceral projection. *Res. Publ. Ass. nerv. ment. Dis.*, 1952, 30, 371–402.

7. Applezweig, M. H., & Moeller, G. Hormonal influences in learning: the pituitary-adrenal system, anxiety, and avoidance learning. ONR project reports NR 154–103 and NR 154–137, 1953.

8. Auld, F., Jr. The effects of TEA on a habit motivated by fear. *J. comp. physiol. Psychol.*, 1951, 44, 565–574.

9. Bailey, P., & Bremer, F. A sensory cortical representation of the vagus nerve. *J. Neurophysiol.*, 1938, 1, 405–412.

10. Bitterman, M. E., Reed, P., & Krauskopf, J. The effect of the duration of the unconditioned stimulus upon conditioning and extinction. *Amer. J. Psychol.*, 1952, 45, 256–262.

11. Brady, J. V. Does tetraethylammonium reduce fear? *J. comp. physiol. Psychol.*, 1953, 46, 307–310.

12. Brady, J. V., Schreiner, L., Geller, I., & Kling, A. Subcortical mechanisms in emotional behavior: the effect of rhinencephalic injury upon the acquisition and retention of a conditioned avoidance response in cats. *J. comp. physiol. Psychol.*, 1954, 47, 179–186.

13. Brogden, W. J. Acquisition and extinction of a conditioned avoidance response in dogs. *J. comp. physiol. Psychol.*, 1949, 42, 296–302.

14. Brown, J. S., & Jacobs, A. The role of fear in the motivation and acquisition of responses. *J. exp. Psychol.*, 1949, 39, 747–759.

15. Brush, F. R., Brush, Elinor S., & Solomon, R. L. Traumatic avoidance learning: the effect of the CS-US interval on acquisition and extinction. Paper read at East. Psychol. Ass., Boston, Mass., April, 1953.

16. Cannon, W. B. *The wisdom of the body.* New York: Norton, 1932.

17. Clark, L. D., Quarton, G. C., Cobb, S., & Bauer, W. Further observations on mental disturbances associated with cortisone and ACTH therapy. *New Eng. J. Med.*, 1953, 249, 178–183.

18. Clark, W. E. Le Gros, & Meyer, M. Relationships between the cerebral cortex and the hypothalamus. *Brit. Med. Bull.*, 1950, 6, 341–344.

19. Cleghorn, R. A. Endocrine influence on personality and behavior. In *The biology of mental health and disease.* New York: Hoeber, 1952. Pp. 265–276.

20. Cobb, S. *Emotions and clinical medicine.* New York: Norton, 1950.

21. Dykman, R. A., & Gantt, W. H. A comparative study of cardiac and motor conditional responses. *Amer. J. Physiol.*, 1951, 167, 780.

22. French, J. D., von Amerongen, F. K., & Magoun, H. W. An activating system in brain stem of monkey. *Arch. Neurol. Psychiat.*, 1952, 68, 577–590.

23. Freud, S. *The problem of anxiety.* New York: Norton, 1936.

24. Fulton, J. F. *Physiology of the nervous system.* (3rd Ed.) New York: Oxford Univer. Press, 1949.

25. Fulton, J., Pribram, K. H., Stevenson, J. A. F., & Wall, P. Interrelations between orbital gyrus, insula, temporal tip, and anterior cingulate. *Trans. Amer. Neurol. Ass.*, 1949, 74, 175–179.

26. Gantt, W. H. *Experimental basis for neurotic behavior.* New York: Hoeber, 1944.

27. Gantt, W. H. Principles of nervous breakdown—schizokinesis and autokinesis. *Ann. N. Y. Acad. Sci.*, 1953, 56, 143–163.

28. Graham, B. F. Neuroendocrine components in the physiological response to stress. *Ann. N. Y. Acad. Sci.*, 1953, 56, 184–199.

29. Greenacre, P. *Trauma, growth, and personality.* New York: Norton, 1952.

30. Gwinn, G. T. The effects of punishment on acts motivated by fear. *J. exp. Psychol.*, 1949, 39, 260–269.

31. Harris, G. W. The hypothalamus and endocrine glands. *Brit. Med. Bull.*, 1950, 6, 345–350.

32. Hebb, D. O. *Organization of behavior.* New York: Wiley, 1949.

33. Hilgard, E. R., & Marquis, D. G. *Conditioning and learning.* New York: D. Appleton-Century, 1940.

34. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.

35. Hume, D. M., & Wittenstein, G. J. The relationship of the hypothalamus to pituitary-adrenocortical function. In J. R. Mote (Ed.), *Proceedings of 1st Clinical ACTH Conference.* Philadelphia: Blakiston, 1950.

36. Jenkins, W. O., & Stanley, J. C. Partial reinforcement: a review and critique. *Psychol. Bull.*, 1950, 47, 193–234.

37. Jersild, A. T., & Holmes, F. B. Methods of overcoming children's fears. *J. Psychol.*, 1935, 1, 75–104.

38. JONES, H. E. The retention of conditioned emotional responses in infancy. *J. genet. Psychol.*, 1930, 37, 485–498.

39. KAMIN, L. J. Traumatic avoidance learning: the effects of CS-US interval with a trace conditioning procedure. *J. comp. physiol. Psychol.*, 1954, 47, 65–72.

40. KIMBLE, G. A. A description of avoidance learning. *Amer. Psychologist*, 1952, 7, 271. (Abstract)

41. KUBIE, L. S. Some implications for psychoanalysis of modern concepts of the organization of the brain. *Psychoanal. Quart.*, 1953, 22, 21–68.

42. LAUER, D. W. Conditioning under complete curarization. *Amer. Psychologist*, 1951, 6, 280. (Abstract)

43. LICHTENSTEIN, P. E. Studies of anxiety: I. The production of a feeding inhibition in dogs. *J. comp. physiol. Psychol.*, 1950, 43, 16–29.

44. LIDDELL, H. S. Conditioned reflex method and experimental neurosis. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* New York: Ronald, 1944. Pp. 389–412.

45. LIDDELL, H. S., ANDERSON, O. D., KOTYUKA, E., & HARTMEN, F. A. Effect of extract of adrenal cortex on experimental neurosis in sheep. *Arch. Neurol. Psychiat.*, 1935, 34, 973–993.

46. LIGHT, J. S., & GANTT, W. H. Essential part of reflex arc for establishment of conditioned reflex; formation of conditioned reflex after exclusion of motor peripheral end. *J. comp. Psychol.*, 1936, 21, 19–36.

47. LINDSLEY, D. B. Emotion. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 473–516.

48. LINDSLEY, D. B. Brain stem influences on spinal motor activity. *Res. Publ. Ass. nerv. Ment. Dis.*, 1952, 30, 174–195.

49. LORENTÉ DE NO, R. Studies on the structure of the cerebral cortex: I. The area entorhinalis. *J. Psychol. Neurol., Lpz.*, 1933, 45, 381–437.

50. LORENTÉ DE NO, R. Studies on the structure of the cerebral cortex: II. Continuation of the study of the ammonic system. *J. Psychol. Neurol., Lpz.*, 1934, 46, 113–177.

51. LOUCKS, R. B. The experimental delimitation of neural structures essential for learning: the attempt to condition striped muscle responses with faradization of the sigmoid gyri. *J. Psychol.*, 1935–36, 1, 5–44.

52. MACLEAN, P. D. Psychosomatic disease and the "visceral brain": recent developments bearing on the Papez theory of emotions. *Psychosom. Med.*, 1949, 11, 338–351.

53. MACLEAN, P. D. Some psychiatric implications of physiological studies on frontotemporal portion of limbic system (visceral brain). *EEG clin. Neurophysiol.*, 1952, 4, 407–418.

54. MAGOUN, H. W. An ascending reticular activating system in the brain stem. *Arch. Neurol. Psychiat.*, 1952, 67, 145–154.

55. MAIER, N. R. F. *Frustration, the study of behavior without a goal.* New York: McGraw-Hill, 1949.

56. MASSERMAN, J. H. *Behavior and neurosis.* Chicago: Univer. of Chicago Press, 1943.

57. MASSERMAN, J. H., & PECHTEL, C. Neuroses in monkeys: a preliminary report of experimental observations. *Ann. N. Y. Acad. Sci.*, 1953, 56, 253–265.

58. MILLER, N. E. Studies of fear as an acquirable drive: I. Fear as motivation and fear-reduction as reinforcement in the learning of new responses. *J. exp. Psychol.*, 1948, 38, 89–101.

59. MILLER, N. E. Learnable drives and rewards. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 435–472.

60. MIRSKY, I. A., MILLER, R., & STEIN, M. Relation of adrenal cortical activity and adaptive behavior. *Psychosom. Med.*, in press.

61. MOWRER, O. H. Anxiety-reduction and learning. *J. exp. Psychol.*, 1940, 27, 407–516.

62. MOWRER, O. H. On the dual nature of learning—a reinterpretation of "conditioning" and "problem-solving." *Harv. educ. Rev.*, Spring, 1947, 102–148.

63. MOWRER, O. H. *Learning theory and personality dynamics.* New York: Ronald, 1950.

64. MOWRER, O. H., & JONES, H. M. Habit strength as a function of the pattern of reinforcement. *J. exp. Psychol.*, 1945, 35, 293–311.

65. MOWRER, O. H., & LAMOREAUX, R. R. Avoidance conditioning and signal duration—a study of secondary motivation and reward. *Psychol. Monogr.*, 1942, 54, No. 5 (Whole No. 247).

66. MOWRER, O. H., & LAMOREAUX, R. R. Fear as an intervening variable in avoidance conditioning. *J. comp. Psychol.*, 1946, 39, 29–50.

67. Papez, J. W. A proposed mechanism of emotion. *Arch. Neurol. Psychiat.*, 1937, **38**, 725.
68. Pavlov, I. P. *Conditioned reflexes.* London: Oxford Univer. Press, 1927.
69. Rome, H. P., & Braceland, F. J. Use of cortisone and ACTH in certain diseases: psychiatric aspects. *Proc. Staff Meet., Mayo Clinic,* 1950, **25**, 495–497.
70. Schlosberg, H. The relationship between success and the laws of conditioning. *Psychol. Rev.*, 1937, **44**, 379–394.
71. Schoenfeld, W. N. An experimental approach to anxiety, escape and avoidance behavior. In P. H. Hoch & J. Zubin (Eds.), *Anxiety.* New York: Grune and Stratton, 1950. Pp. 70–99.
72. Schreiner, L., & Kling, A. Behavioral changes following rhinencephalic injury in the cat. *J. Neurophysiol.*, in press.
73. Selye, Hans. *The physiology and pathology of exposure to stress.* Montreal: Acta, Inc., 1950.
74. Sheffield, F. D. Avoidance training and the contiguity principle. *J. comp. physiol. Psychol.*, 1948, **41**, 165–177.
75. Sheffield, F. D., & Temmer, H. W. Relative resistance to extinction of escape training and avoidance training. *J. exp. Psychol.*, 1950, **40**, 287–297.
76. Sheffield, Virginia F. Extinction as a function of partial reinforcement and distribution of practice. *J. exp. Psychol.*, 1949, **39**, 511–525.
77. Skinner, B. F. Two types of conditioned reflex and a pseudo type. *J. gen. Psychol.*, 1935, **12**, 66–77.
78. Solomon, R. L., Kamin, L. J., & Wynne, L. C. Traumatic avoidance learning: the outcomes of several extinction procedures with dogs. *J. abnorm. soc. Psychol.*, 1953, **48**, 291–302.
79. Solomon, R. L., & Wynne, L. C. Avoidance conditioning in normal dogs and in dogs deprived of normal autonomic functioning. *Amer. Psychologist*, 1950, **5**, 264. (Abstract)
80. Solomon, R. L., & Wynne, L. C. Traumatic avoidance learning: acquisition in normal dogs. *Psychol. Monogr.*, 1953, **67**, No. 4 (Whole No. 354).
81. Starzl, T. E., & Whitlock, D. G. Diffuse thalamic projection system in monkey. *J. Neurophysiol.*, 1952, **15**, 449–468.
82. Stern, M. M. Anxiety, trauma, and shock. *Psychoanal. Quart.*, 1951, **20**, 179–203.
83. Trethowan, W. H., & Cobb, S. Neuropsychiatric aspects of Cushing's syndrome. *Arch. Neurol. Psychiat.*, 1952, **67**, 283–309.
84. Wall, P. D., & Davis, G. D. Three cerebral cortical systems affecting autonomic function. *J. Neurophysiol.*, 1951, **14**, 507–517.
85. Ward, A. A., Jr. The anterior cingulate gyrus and personality. *Res. Publ. Ass. nerv. ment. Dis.*, 1948, **27**, 438–445.
86. White, R. W. *The abnormal personality.* New York: Ronald, 1948.
87. Wynne, L. C., & Solomon, R. L. Traumatic avoidance learning: acquisition and extinction in dogs deprived of normal peripheral autonomic function. *Genet. Psychol. Monogr.*, in press.
88. Zanchetti, A., Wang, S. C., & Moruzzi, G. The effect of vagal afferent stimulation on the EEG pattern of the cat. *EEG clin. Neurophysiol.*, 1952, **4**, 357–361.
89. Zener, K., & McCurdy, H. G. Analysis of motivational factors in conditioned behavior: I. The differential effect of changes in hunger upon conditioned, unconditioned, and spontaneous salivary secretion. *J. Psychol.*, 1939, **8**, 321–350.

(Received February 4, 1954)

Psychological Review
Vol. 61, No. 6, 1954

THE PSYCHOANALYTIC THEORY OF CONFLICT: STRUCTURE AND METHODOLOGY [1]

ALLEN T. DITTMANN [2] AND HAROLD L. RAUSH
University of Michigan

Although some criticisms of psychoanalytic theory can be laid at the door of the personal motivations and unconscious resistances of the critics, there is something to the view that workers in the field do not have available to them any concise and complete statement of the structure of psychoanalytic theory. Accordingly, there is a large group who mistake the *content* of psychoanalytic findings for the *structure* of psychoanalytic theory, and believe, for example, that the stages of psychosexual development and the standard list of defense mechanisms are coterminous with the theory itself. Caught in this confusion, many investigators have contented themselves with attempting to "prove" or "disprove" "psychoanalysis" on the basis of studies of certain content elements of the theory. There can be no doubt that the investigation of isolated aspects of psychoanalytic theory may furnish us with factual information and with broad differentiations. We believe, however, that only through understanding and utilizing the structural interrelations among the concepts of the theory can we make use of the real power and research potential which the theory may offer, or enter into legitimate conceptual refinements and revisions.

The purpose of this paper is, therefore, to outline these structural interrelations. We shall concern ourselves with the psychological *processes* that psychoanalysis posits as necessary for the understanding of certain behaviors. In view of our limited purpose, it will be necessary to omit or reduce discussion of many relevant aspects and complexities of psychoanalytic theory. Thus we shall not deal, for example, with the theories of psychosexual stages, system structure (id, ego, superego), intrasystem problems, differentiation of impulses (sex and aggression), primary and secondary processes, or with many clinical and technical problems. Our presentation is drawn from Freudian psychoanalytic theory; after presenting the theoretical outline, we shall examine the research methodology from which this viewpoint developed, and discuss some research directions toward which psychoanalysis points.

Sources of Data for the Theory

In order to understand the structure of the theory and its sometimes confusing syntax, it is first necessary to look at the sources of data used in its development. Many sources have been drawn upon, from ethnological field reports to children's drawings, but all of them have been used chiefly as collaborative evidence for an already growing theory: the primary source of data has always been the productions of psychoneurotic patients undergoing therapeutic analyses. The fact that the data supplied by patients did not fit contemporary notions led Freud to break away from his late nineteenth century coworkers and to start the development of psychoanalytic therapy. Similarly, the fact that the data of the treatment

[1] The writers wish to express appreciation to Edward S. Bordin, Donald G. Marquis, and Frederick Wyatt for their critical reading of the manuscript. The responsibility for the views expressed is, however, that of the two authors, who shared equally in the development and writing of the paper.

[2] Now at National Institute of Mental Health, Bethesda, Maryland.

situation did not fit his own early formulations led Freud and his new coworkers to make every major change, expansion, and development that the theory has undergone up to the present time.

Psychoanalysis developed in a clinical setting as a therapeutic measure, at first particularly for hysterical patients; it was primarily a psychology of the unconscious (13, p. 128). One of the first clinical facts that Freud recognized was that a psychology not based on unconscious sources of motivation was inadequate either for understanding or for treating neurotic patients. The method of free association was gradually evolved as the most economical method for clarifying unconscious sources of behavior and their effects. The basic assumption of this method is that free associations are indeed not "free," but are rather highly determined by unconscious motivations. The fundamental rule "to say everything that comes into your mind, whether it seems important or unimportant, relevant or irrelevant," is designed to minimize the influence of consciously directed thought processes, and to maximize the influence of unconscious factors. Other aspects of the analytic situation as it is usually constituted—the reclining position of the patient, the subdued atmosphere of the consulting room, the position of the analyst out of direct sight of the patient—are similarly designed to maximize the influence of unconscious factors in the patient's "free" associations.

Derivatives as behavioral resultants of unconscious conflict. Under these traditionally analytic conditions, many, though not necessarily all, of the patient's productions in the analytic hour are derivatives of his unconscious conflicts. That is, these productions are representatives of the same conflicts that have led him into the difficulties for which he has sought treatment. Precisely what conflict the patient's productions represent is at first unknown to both patient and analyst. Even the fact that his productions do represent conflict may be unknown to or of no concern to the patient, but this fact is the major assumption under which the analyst operates. Were the associations (or sometimes the lack of them) not representative of conflict, and were they not *disguised* representatives of *unconscious* conflict (that is, derivatives), the patient would have little need or motivation for treatment—he could deal with reality as adequately as the situation allowed, and with a minimal degree of discomfort.

Let us consider a clinical example of the production of derivatives, an example which we may use for further discussion of derivatives and of their importance to the structure of psychoanalytic theory. A student has difficulty in submitting papers on time. While he feels that he is capable of "A" work, he usually manages to get an A − or a B +. Problems such as this are so common in student populations that to speak of them as derivatives of unconscious conflicts is at this point foolhardy. But let us go a step further. When our student gets an A, we are surprised to learn that he feels vaguely dissatisfied: the instructor has misunderstood him, has praised the wrong things. What is more, our student now feels that the ideas are no longer his, that he has given in to external pressure and has let his ideas and thoughts be taken away from him. Yet at the same time he does want to express these ideas and wants them to be appreciated. His communications to the therapist follow the same line. As soon as he has communicated a thought, he feels dissatisfied and tries in various ways to nullify what he has said. He feels that the thoughts he has expressed are no longer his, but are rather the property of the therapist. He blocks at a point where he is discussing his difficulties in communicating to the therapist, and the latter inquires as to the thought at the point of blocking. The patient embarrassedly confesses having thought of feces, but disclaims this as an irrelevant, nasty thought, and can go no further with the subject. We continue to discover similar patterns in his other interpersonal relationships: we find that he wants to give himself up completely in an all-encompassing intimacy with his friends, surrendering his ideas and his will, yet he finds himself in sudden bouts of temper and resentment, fighting in strange ways to maintain his integrity. He sees women clutching at him and trying in devious ways to trap him and get something out of him. Sexual expression and sexual curiosity he feels are signs of his vulnerability and must be denied. At times he fantasies being overwhelmed and raped by one more powerful than he.

THE THEORETICAL STRUCTURE OF PSYCHOANALYSIS

The derivative as compromise behavior. In clinical examples such as the one cited above, the therapeutic situation has allowed for intensive, long-time exploration of the meaning of the derivatives in terms of conflicts underlying them. We have the opportunity to learn how such seemingly superficial behaviors as not submitting term papers on time are ramifications of early childhood experiences. We may, however, observe the formation of derivatives in other situations, too, although the opportunities for exploring their sources are usually not so available to us. Dreams, slips of the tongue, symptoms are examples. All of these phenomena are compromises between opposing forces which make up unconscious conflicts within individuals. The behavior involved in the compromise always expresses in a disguised form some impulse which would be otherwise unacceptable. The disguise itself functions so as to allow expression of the impulse in such a way that the anxiety which would follow direct expression of the impulse is not so likely to be aroused. The processes by which the disguises are accomplished are known as the defensive functions of the ego. Thus, the very naming of a bit of behavior as a derivative implies a complex theoretical structure: an impulse striving for expression, the direct expression of which would involve anxiety, and defensive maneuvers to allow some expression with minimal anxiety.[3]

Impulse

Let us begin our analysis of this theoretical structure with a consideration of impulses. The exact nature of basic impulses is not known, and the issue of their explanation is a confused one. Freud, himself, seems always to have been dissatisfied with his formulations in this area, and was careful to point out the tentative state of his conceptualizations.[4]

[3] Psychoanalytic writers using the term "derivative," generally speak of "derivatives of unconscious impulses." These refer, as above, to distorted forms of impulse expression which occur when direct impulse expression is impossible because of anxiety. The context of the writings makes it clear that the distortion process is as important as the expression, and that derivatives may be considered in terms of their degree of distortion of the original impulse. Since the distortion is a function of defensive processes, the derivative always represents both impulse and defense (6, p. 57). We believe, then, that it is more consistent to speak of a derivative as the resultant of an impulse-anxiety-defense conflict.

[4] One such remark was made in the *New Introductory Lectures on Psycho-analysis:* "The theory of the instincts is, as it were, our mythology" (15, p. 131). The historical development and problems associated with this theory are discussed by Bibring (2).

We use the English word "impulse" here for

We shall tentatively define impulse as a physiologically based tension state that sets behavior in motion. Freud (11) describes impulses as having (*a*) impetus, which is the motor element of drive, the force it represents; (*b*) aim, which is, in general, discharge of the energy or satisfaction; (*c*) object, which is the instrument through which satisfaction may be attained; and (*d*) source, which is the physicochemical state of the organism.[5]

Impetus. We may consider impetus as representing the quantitative strength of an impulse, a strength which varies from time to time. It decreases when certain specific behavior patterns occur in conjunction with specific environmental opportunities; it increases either where the specific behavior pattern fails to occur or where the environment fails to present the appropriate opportunities. In words more generally used by psychoanalysts, the impetus of an impulse decreases with satisfaction and increases with inhibition or deprivation.

Aim. While it is true that the aim of impulses is always toward discharge, toward an eventual lowering of tension level, the process may be inhibited or deflected, and the modes of achieving this aim may be varied. At times Freud seems to use the word "aim" to include differing modes of impulse satisfaction. An example of this confusion may be found in his discussion of the "change of aim" of an impulse from active to passive, as scoptophilia to exhibitionism, and sadism to masochism (11, p. 72). Here the phrase "change of aim" should more properly read "change of mode for achieving aim," as Freud wrote elsewhere in the same paper (11, p. 69).

It should be noted that the immediate satisfaction of some impulses, such as respiration, is imperative to the life of the organism, whereas for other impulses, greater and perhaps indefinite delay can be tolerated. Similarly modes of satisfaction are more variable for some impulses than for others.

Object. Impulses vary in the range of objects that may serve to reduce their impetus. Respiratory needs, for example, can be satisfied only by breathing air. Sexual impulses, on the other hand, can be satisfied by a wide variety of objects and by various modes. The impulses most relevant to psychology are those most amenable to variation in mode of expression and in object. These impulses are important because they are the ones most subject to environmental influences and hence can become the foci of interpersonal conflict.

Source. Freud's belief in physicochemical states as the sources of all impulses forms the basis of his biological viewpoint. He even hoped for a state of knowledge which would allow classification of impulses on the basis of their sources (11). We are still unable to fulfill these hopes, although the time may yet come when we will be able to do so.

Psychological Representations of Impulse

In talking about impulses as physiological states, it is of the utmost importance to recognize that these states themselves are not psychological phenomena. The psychological phenomena are, rather, the tensions experienced by the individual, tensions which carry with them, to a greater or lesser degree, content of associations with other ex-

the German *Trieb*, which has usually been translated "instinct," and occasionally "drive" (7, p. 12).

[5] Freud appears to have dropped the term impetus in his later writings, though equivalent concepts appear in his discussion of strength of instincts (16); Fenichel (7) discusses aim, object, and source as characteristics of instincts, but Sterba (25) includes the concept of impetus in his discussion, though he calls it "drive."

periences. These tensions are not necessarily related in a one-to-one fashion with fluctuations in the physiological state of the organism. Thus, operationally it is necessary to distinguish between the physiological state that results from food deprivation, for example, and the experience of the resulting tensions. This psychological hunger varies not only as a function of the physiological state, but also as a function of such conditions as age and previous experience of the individual. These psychological concomitants of physiological states are what Freud called "impulse presentations" (12). In most clinical writing the concepts of impulse and impulse presentation have unfortunately been confused, so that the term impulse has come to be used as a standard expression for impulse presentation. Consequently, in psychoanalytic literature, impulse variously refers to: (*a*) an energic, "driving," physiological force; (*b*) the psychological representations of such forces; and (*c*) residues of experiences with such forces.

We may illustrate this confusion by considering again the case of the student cited above. For this man, current sexual tensions, when they arise, are accompanied by a multiplicity of contents. It may be possible to speak of each of these contents as independent impulses or as fusions of them, e.g., scoptophilic, passive-dependent, anal, homosexual, and other impulses. However, to the extent that we believe these factors to be modifiable, an inconsistency is involved in speaking of them as impulses in the sense of independent physiological tension states. We should rather speak of such tendencies as impulse presentations concomitant with a physiological imbalance, which in this case is sexual tension. It seems reasonable to suppose that our patient's twistings and turnings with respect to sexual impulses are not a function of struggle between impulses, but rather that they are a function of memory traces (cf. 9, ch. VII) resulting from his own unique experiences. These experiences have influenced the course of handling states of physiological imbalance.

In the case of another patient, psychological tensions were always so high that no one physiological state could be differentiated from any other. She stated, for example, that she never "felt hungry," but ate when her watch pointed to the traditional times for meals. In the course of treatment she came to know periods of relaxation during which specific tensions could be felt. Impulse and impulse presentation were remarkably separate from each other for this patient.

Anxiety

Primary anxiety or panic. For any individual, experiences may have been favorable or unfavorable to the expression of a given impulse, that is, to the correct interpretation of a physiological imbalance and to behavior that will lead to the reduction of this imbalance. Where experience is favorable, the individual gradually develops techniques for achieving impulse satisfaction, and evaluations of the appropriate times and places for impulse satisfaction. Such learnings may be made simple or complex, either by society or by parents as particular representatives of society. As long as society provides channels for discharge and adequate rewards for approved modes of discharge, learning is a continuously modifying process. That is, experience becomes checked and modified by new experience.

But in the case of our patient, we may note that he avoids those situations which are appropriate to impulse expression, and that he misinterprets both the external world and what we infer to be his physiological needs. He is furthermore strikingly persistent in his misinterpretations, despite the seeming discomfort that these misinterpretations bring. The evidence he presents leads us to believe that experiences in his life had the peculiar effect of preventing him from profiting by new experiences. How does psychoanalytic theory account for such a state of affairs?

The expression of any impulse or of all impulses may be accompanied by intense threat in the early experience of an individual. This threat consists of the overwhelming of the psychic apparatus by tensions attendant upon physiological imbalances that the infant cannot satisfy by himself. These imbalances may be the result of the normal periodic rise of needs which parental figures have not yet satisfied at the moment, or they may result from external causes such as parental treatment, disease, and the like. Where parental treatment in relation to impulse expression leads to such overwhelming threat, the threat is likely to be a repeated experience, either because of institutionalized child-rearing practices or because of the individual parent's persistent needs and fears. The response of the infant to such trauma has been called primary anxiety (7, pp. 132 ff.).

In its theory of genetic development, psychoanalysis proposes certain physiological and psychological stages. For the purposes of the present paper it is not necessary to go into the details of these stages; they have been described adequately elsewhere (3, 5, 7). What is relevant here is that these stages are associated with certain impulses and threats to their expression, and consequently have potentiality for becoming focal points for primary anxiety.[6] The panic experience of primary anxiety is so painful that the psychological representation of the impulse comes to be regarded as dangerous. Similarly, affects and images which are concomitant with the experience of primary anxiety often come to be perceived as dangerous. According to the theory, such experiences and their psychological connections tend to undergo repression and are retained as unconscious memory traces.

Secondary anxiety and memory traces. Having experienced primary anxiety, the infant is sensitized to those events which might rearouse panic. These events may take two forms: (*a*) from the impulse side, any increase in the impetus of the specific impulse; (*b*) from the side of the external world, any situation which is provocative of impulse expression or which is similar to the events surrounding the earlier experience of panic. In either form, such events bring about expectations on the part of the individual that a panic of the same intensity as the original one will arise. The individual is consequently always on the alert for cues from within and from without that will signal an impending danger similar to that experienced in the past. This signaling, which is an attenuation of the earlier panic experience, is known as anxiety. In Freud's words, "Anxiety . . . is the expectation of the trauma on the one hand, and on the other, an attenuated repetition of it" (14, p. 114).

[6] The question of the first experience of anxiety and its effects on later experiences of anxiety has been one of controversy to the point of forming separate "schools" of psychoanalysis: the intense pain of hunger in the first few weeks of life, the shock of birth, or prenatal experiences have all been posited as the primary experience. All writers agree, however, that no matter what the initial circumstances might have been for an individual, later experiences of anxiety are in some sense referred to earlier ones.

The individual's problem, then, is to use the anxiety signal to become active and thus prevent the attenuated panic from developing into a full-blown one.

Since impulses continue to arise and tend toward motor expression, there is an ever recurrent pattern in the life of the individual: anxiety develops and attempts are instituted to prevent anxiety from growing into panic. The theory seems to imply (7, p. 143 f.) that under such circumstances new experiences do not function so as to modify infantile patterns of behavior and perception, but rather become assimilated into the memory traces connected with the impulse.[7] We must assume that any expression or recognition of the nature of the impulse or of the memory trace would lead to a recurrence of the original panic. We shall have more to say about this, but let us turn first to the role of the current situation.

The Role of the Current Situation [8]

At any given time, the momentary strength of an impulse is a function of two factors: (*a*) from within, it is related to impetus, which is a function of the state of deprivation of the organism, as described above; (*b*) from without, it is related to the "provocativeness" of situations for the expression of the impulse. The provocativeness of various situations for impulse expression is a difficult psychological problem. As we have pointed out in the above discussion of object, there may be cultural variations in the range of appropriate situations for impulse expression. Beyond these variations, there are modifications as a function of individual training. Nonetheless, it is necessary to assume that biologically some impulse-situation configurations are more "natural" than others, and that cultural modifications must confine themselves within certain limits. In the ecology of events in the environment, more and less provocative situations will occur for any given impulse. The greater the provocativeness of the situation, at any given degree of impetus, the greater will be the momentary impulse strength. As momentary impulse strength increases, the greater will be the tendency toward direct impulse expression, and the greater the energy required to inhibit direct expression.

Impulse expression does not occur in vacuo, but rather requires commerce with the environment. Thus people will seek out situations for impulse satisfaction when impetus increases. Where "natural" situations are unavailable, or where they are too anxiety provoking because of memory traces related to the impulse, then substitutions are found for the more natural ones.

Defense

There is a host of consequences to the inhibition of impulse expression. First, the impetus of the impulse will increase. With this increase in impetus there is an increasing likelihood of rearousal of panic if the latter is latent in the memory trace of the impulse. Second, the increased impetus demands increased effort on the part of the individual to prevent panic; in order to maintain equilibrium within the system,

[7] The theory is not clear on whether this course of events depends on the quantitative level of momentary anxiety and/or the intensity of the original panic, or whether it depends on the qualitative presence or absence of anxiety. The question is a crucial one for "ego psychology" (17).

[8] The current situation is rarely considered as a systematic variable in psychoanalytic theory. Its role is, however, clearly implied in many psychoanalytic writings. See, for example, the discussion on the role of the experiences of the previous day in the process of dream formation (9), and the effects of the current situation in relation to psychotherapy (16). The concept of momentary impulse strength does not appear in psychoanalytic writings, but seems useful to us.

energy must be expended for the prevention of panic. The energy bound up in this process, which has been called countercathexis, is not available for other organismic functions (7, p. 141 f.).

Let us consider the case where there is increased impetus and a "provocative" environmental situation. Let us consider, furthermore, that the experience of the individual (in the form of memory traces in the impulse presentation) has been such that greater increase in the momentary impulse strength will arouse panic. Anxiety signals this possibility of panic, and becomes a cue for the organism to institute defensive measures against panic. Such defensiveness may take many forms.[9] In some of these forms anxiety is allayed by avoiding recognition of or by distortion of the impulse; other forms function so as to avoid recognition of the provocativeness of the situation or by distortion of the situation. All defensive maneuvers to some extent and at some level avoid or distort the relationship between impulse and situation. When, as a function of anxiety, such distortions are necessary, new situations do not serve to alter perceptions or behavior in relation to impulses and their satisfaction.[10]

Behavioral Resultants: Derivatives

Having come by the devious route through discussion of impulse, memory trace, situation, anxiety, and defense, we are finally in a position to consider the behavior that results from the interplay of all these factors: the derivative of an unconscious conflict. In psychoanalysis behavior is seen either as rational, reality-oriented, appropriate, and goal-directed, or as irrational, irreality-oriented, inappropriate, and directed toward anxiety avoidance. Our judgments on these issues leave something to be wanted. Nevertheless, we can make approximations. Where behavior is consciously directed toward a goal; where it is integrated with other behaviors; where it is capable of modification with the acquisition of information; where interruption of the behavior leads to a search for other means-end objects or rational subgoals or substitute goals; and where completion of a behavior sequence results either in satisfaction or a modification of behavior, then we deem behavior to be appropriate. On the other hand, where an individual's behavior appears to be unintegrated and inconsistent with his other behaviors; where it is incapable of modification in the light of situational variants; where interruption leads to anxiety; and where completion of a behavior sequence leads neither to satisfaction nor to a modification of the sequence, we are led to believe such behavior to be a derivative of an unconscious conflict.[11]

Let us turn once more to our clinical example for concrete illustration of the abstractions we have been discussing. We do not know the specific natures of the impulses we are dealing with. We have evidence in the case of the student that there is a common impulse presentation that arises when he ap-

[9] The classical classification of defenses may be found in Anna Freud's *The Ego and the Mechanisms of Defence* (8).

[10] We have left the concepts of guilt and shame out of our discussion. These are often considered to be forms (with certain unique characteristics) that anxiety may take (7). These also lead to inhibitions and defenses.

[11] These listings are not to be construed as complete descriptions of "rational" and "irrational" behavior. Psychoanalytic theory has for the most part been concerned with irrational behavior. Freud himself was much interested in understanding rational behavior (cf. 20), and there are attempts currently at a psychoanalytic psychology of rational behavior (5; 17; 18; 22; 23, pp. 689–731). Recent psychoanalytic writings (cf. 17) imply that the distinction between rational and irrational processes and behavior is not as clear cut as we have suggested here.

pears to be seeking, or when situations call for responses of love, affection, sex, warmth in relationships. Such affects in the patient's past have been associated with panic. When in his early life wishes relating to these affects appeared, the patient's mother reacted with anxiety and treated him in such a way that psychological and physiological equilibrium could not be re-established. When he is currently tempted (either by recurrent impulse increases or by provocative situations) toward expressions in these areas, the memory trace of panic gives rise to anxiety. This the patient tries to allay by projection—he perceives others as making demands on him, and then complies with these perceived demands, but always in such a way as to create the impression that he is not doing his best. He thus leads people to believe that there are vast areas of untapped potentiality in him (as indeed there are), and they seek to guide and encourage him. In this manner he obtains some measure of love, while at the same time he is able to deny that he himself is seeking it. The behavior, as we see, exhibits both impulsive aspects (for example, seeking love) and defensive aspects (for example, projection).

By its very nature as the resultant of an unconscious impulse-defense conflict, a derivative does not provide true impulse satisfaction. At best, it lowers tension level momentarily.[12] Because of its unreality aspects, the production of derivatives tends to become progressively complicated, as for example when our patient feels resentment toward the people upon whom he has projected his own demands. The resentment against external demands has in its turn memory traces associated with it, in terms of the threat of loss of love. Anxiety arises when the expression of resentment is imminent, and further defensive measures are instituted by denial of affect. The resulting derivative behavior is often a superficial politeness which carries a tinge of supercilious condescension. The progressive development of this process has been referred to as the layering of defenses (6, 7, 24). Thus, the less satisfying the derivative: the greater the likelihood for more complicated defensive structure; the greater the amount of energy expended in maintaining the structure itself; the more numerous the areas of living permeated by pathology; and the less energy left over for other satisfactions.

THE METHODOLOGY OF PSYCHOANALYTIC RESEARCH

Psychoanalysis developed rather independently of the tradition of experimental psychology. Psychology, in its academic setting and in its consciousness of itself as a new science, has consistently had the inclination and has found the time to examine the status of its methods. Psychoanalysis, on the other hand, developed in a clinical setting. The training of its workers, the daily exigencies of its practice, and the impelling force of the fruitfulness of its discoveries, have militated against a careful perusal of the assumptions and methods of its research. Though there are some exceptions, it is only recently that psychoanalytic writers have begun to turn from clinical problems back to the more general theoretical issues which so concerned Freud and his early coworkers (cf. 20). If we give any credence to the discoveries of psychoanalysis, it is to the interest of not only clinical workers, but to all those concerned with the behavioral sciences, to look seriously at its methodology. For it would seem specious to accept its discoveries, even if only as a fruitful

[12] How a response which does not lead to direct physiological satisfaction of an impulse can lower tension level at all is unclear in the theory.

source of hypotheses, and simultaneously dismiss the methods of making these discoveries. In this section we can only touch on a few of the methodological problems that grow out of our earlier discussion, and on some relationships between psychoanalytic and experimental research.

The Research Method

In common with other psychological systems, the observables for psychoanalytic theory are stimulus situations or the descriptions of stimulus situations and behavior or behavior descriptions. All other concepts are either hypothetical constructs or intervening variables. In order to point out the direction which psychoanalytic research takes within this framework, let us consider the nature of the therapist's activity as experimenter and empirical observer.

The patient describes, in the course of therapy, a wide variety of situations including the therapeutic one. Similarly, he describes and manifests a wide variety of behaviors. One assumption that the therapist works with in forming his hypotheses is that all situations that involve the same behavior are functionally equivalent. For example, if a patient responds to a variety of self-described situations or to the therapeutic situations with anger (or it may be twitching or weeping or withdrawal or descriptions of these), the therapist then forms a hypothesis which would relate these situations. By further observations, questions, and interpretations, he proceeds to check, modify, and extend his hypothesis. Thus he may observe that the behavior (or again, the patient's description of his behavior) of anger follows situations (or descriptions of situations) where affectionate feelings are aroused in the patient. He hypothesizes this relationship to be invariant. If the patient then expresses anger toward him, the therapist checks with the patient as to whether an unverbalized feeling of affection toward the therapist did not precede the angry outburst. Where his hypothesis fails to be confirmed, it is either completely or in some respect in error. Thus the therapist may, for example, modify the hypothesis to the form that there is an invariant relationship for this patient between affection-producing situations and anger in relationships with women but not with men, or through further checking, the modification may be of invariance in the case of women who behave in a seductive manner, but not toward those who behave in a nonseductive manner and not toward men. On the other hand, a more extensive revision of the hypothesis may be necessary. Such is the case, for example, should the patient become angry not only in one of the above described situations, but also, let us say, when the therapist fails to keep an appointment or is inattentive. We are now in the position of maintaining that there is a realm of situations (S_1, S_2 . . .) which invariantly leads to response R_1, and that there is at least one other realm of situations (S_A, S_B . . . , and perhaps S_α, S_β . . .) which also leads to response R_1, but that the former realm of situations is not equivalent to the latter realm. If we are to maintain our working assumption of the functional equivalence of all situations that involve the same behavior, then we must modify the hypothesis on the behavior side. Thus we may say that one realm of situations will produce response R_1 in a context of responses R_2, R_3 . . . , and that another realm of situations will produce response R_1 in a context of responses R_A, R_B . . . . Our definition of behavior now considers a series of responses.[13]

[13] The above discussion has dealt with the assumption of the psychological equivalence of objective situations leading to the same be-

The procedure we have discussed may at first glance seem to be overwhelmingly complicated. More than that, it may seem to the reader to be lacking in any sort of scientific rigor. Even were we to assume the objectivity of the therapist as observer, the pitfalls of the process are obvious. On the one hand our hypothesis can become so modified in terms of an increasingly refined description, that it serves to describe only a single event. As such, the hypothesis may be true but scientifically useless. On the other hand, it may become so broad in its extensions that it says nothing more than that people will demonstrate behavior in response to environmental events. As such, this may be true, but again scientifically useless. The problem in using such a procedure is one of achieving parsimony and specificity of prediction and is common to all science.

Unlike many of our more nomothetic investigations, the procedure is based on strict scientific determinism. It is similar, perhaps, to the chemical analysis of an unknown compound. Various hypotheses are brought forward, various reagents and tests are employed, hypotheses are modified and refined, and the goal is an accounting for all of the data with a minimal number of concepts. The table of elements, so to speak, and an accounting of their interactions are indeed unprecise in psychoanalytic theory, but there are guideposts for the observer. These guideposts are to be found in the theory discussed earlier in this paper, and in the genetic and structural theories of psychoanalysis. Thus the psychoanalytic observer can gauge to some extent via a succession of temporal events: the degree and type of defensiveness involved in a given derivative behavior, the similarity among behaviors, the points at which anxiety and defenses against anxiety appear, the kinds of impulses which may be broadly involved. From the patient's responses he can form tentative maps of the traces involved. From the genetic theory of such traces he derives hypotheses, and the empirical testing of these hypotheses leads to revision and refinement of the tentative mappings of traces. The investigation then extends to other patients and to the formulation of broader hypotheses. Ideally, at least, there is a constant interaction between theory and data.[14]

The Individual Case as a Source of Data

As the above discussion implies, psychoanalysis is based upon research that involves the long-term study of individual cases. At first blush it would seem that the case study could not be considered research data in the usual sense, for even in these days of small sample techniques, our methods are not capable of handling samples as small as one. If we look at the individual case from the standpoint of derivative theory, however, the perspective changes markedly. Here we are interested in the relationships among impulse, memory trace, situation, anxiety, and defense, and within any one case repeated and stable patterns of these factors may be found.

Problems of the psychoanalytic case study. General issues regarding case study methods have been discussed in the literature (1, 21). We wish, how-

havior. Similarly, a series of hypotheses can in much the same way be based on the assumption that different behaviors following the same objective situation are functionally equivalent. Since we are speaking of temporal sequences, the word "same" with respect to behavior or situation always involves a proximate judgment of "roughly the same." We shall touch on this issue below. Both assumptions are utilized in the temporal process of psychotherapy.

[14] A further discussion of the historical process of this interaction may be found in (19).

ever, to consider briefly some problems pertaining to the psychoanalytic case study as a method of developing and testing theories of behavior. There is one criticism frequently leveled against the use of such data for research purposes. First, since he has devoted so much time and energy to his training, the analyst looks for evidence to corroborate his favorite theory or subtheory, and he will naturally be able to find it in the mass of data which the patient produces. Worse still, since his beliefs are so strong, he will influence the patient, either consciously or unconsciously, to produce corroborative evidence. These are points on which Freud admonished his followers (and also himself) over 40 years ago (10, pp. 326–328), and they are still worth thinking about.

Freud's answer was that if the analyst is able to listen *freely* to the material presented by the patient (and he defined this sort of listening as a task corresponding to the patient's free association—and as full of as many pitfalls), then he will obtain free and uncontaminated data upon which to build and test theories. Freud specifically warns the analyst regarding research biases which may affect the ability to listen freely. The therapeutic analysis of the analyst is an attempt to guarantee free listening for both therapeutic and research purposes, but no one would claim that this guarantee has always worked out perfectly.

A further criticism is that the discovery of lawful regularities in the behavior of one individual is no indication that the same regularities will be found in another individual. Only further study can decide to what extent the functions described are universally relevant. This limitation, however, in no way affects the "scientific" legitimacy of the single case study for the formulation of psychological laws. We shall have more to say of this below.

The Units of Behavior

We have been speaking of associations, dreams, parapraxes, symptoms, and the like as the raw data of psychoanalytic research—in short, the behaviors that make up the "material" with which the psychoanalyst works. We shall now examine more closely the nature of these behaviors as research data.

Superficial consideration might lead one to think that the raw materials of psychoanalytic research are simple behaviors. What makes things more complicated, however, is that psychoanalysts are not interested in behaviors per se, but rather in what these behaviors *mean* for the patient. From the psychoanalytic viewpoint any isolated behavior is likely to be meaningless because it does not contain enough information to work with. The method that has been evolved is that of grouping behaviors into sequences and using the sequences themselves as basic data. Let us take a clinical example. A patient has headaches during her therapeutic sessions. The headaches are so severe that she cannot pay attention to anything else—she forgets what she has been discussing before, and she can hardly hear what the therapist is saying. Confronted solely with this behavior, neither the therapist nor the researcher can have much to say. If, however, one looks at the sequence of situations and behaviors in which headaches appear as a single behavioral event, the specific behavior takes on meaning. In one interview the patient talks about a man in whom she is interested, *then* gets a headache, *then* becomes angry that she cannot have a good time with people; in another interview she has physical sensations of warmth, feels light as though a weight had been lifted from her shoulders, *then* gets a headache, *then* berates herself for not being able to talk with the therapist about problems which are frightening to her. In these and other instances of the sequence, the mood changes from mild euphoria and excitement in the first phase, to pain in the second, to depression in the third. As such sequences cumulate and are consistent in their course, and as further material is added, we gain increasing evidence that in this case the headache serves at least two functions, one to distract the patient from thinking about things which may provoke anxiety, and another (and at a deeper level) to punish her for thinking about forbidden things.

As we gather data on different patients who suffer from headaches, we may develop any of a number of hypotheses. One of these may be, for example, that headaches are a reflection of aggression turned inward. In testing such hypotheses we are no longer directly engaged in studying the processes involved in the occurrence of a specific behavior. We have in effect simply translated from behavior to impulse.[15] The question of the adequacy of this or any other symbolic translation is an empirical one. Its confirmation depends upon, among other things, an adequate sampling of a given class of people (in this case, those who have headaches). Tests of such hypotheses, while they are important to the study of the etiology of symptoms, tackle other problems in addition to those posed by the theory of derivatives. The relationship between behavior sequence and unconscious conflict, as exemplified by the above patient with headaches, may hold for one person only or for all people with a given symptom. Within the general theory, it is conceivable, for example, that headaches bear a relationship with aggression turned inward for one patient, a relationship with libidinization of visual activities for another patient, and an organic factor in a third. The problem is a legitimate empirical one, but only when intermediate steps between impulse and derivative are delineated will its investigation be crucial for psychoanalytic theory. Thus, for example, the fact that temperature elevation can be "caused" by any of a number of disease organisms, or can occur even in the absence of such organisms, does not "disprove" germ theory.[16]

Problems of behavior sequences. We turn now to some of the difficulties involved in studying sequences of behavior rather than isolated units of behavior. First, the very complexity of such sequences makes for research difficulties. While observational techniques may be developed with respectable degrees of reliability for identifying a single behavior, any unreliability that remains may be compounded if more than one behavior is studied at the same time. The reliability of judgment of the sequence itself, then, cannot be very high in the present state of the field. This difficulty is made even greater because many of the important sequences involve interaction between two or more people. To the extent that it provides

[15] The symbolic translation and the relationship to data are even more obscure when we select a group of subjects, all of whom have headaches, for studying the relationship between aggression turned inward and a third factor.

[16] A more directly psychoanalytic example is Erikson's (4) study of the Yurok, a tribe in which there is very compulsive concern with holding onto money. Rather than translating from concern with money to concern with feces, Erikson found the mode of "holding onto" associated with oral rather than anal training in Yurok childhood. Erikson is cognizant of the fact that in using a different cultural sample he has not refuted findings in Western European cultures, but rather he has extended the applicability of psychoanalytic theory.

multiple samples of sequences, the long period of time involved in the psychoanalytic treatment situation does help in the matter of practical reliability. But even here, as Freud (16) pointed out, certain situations and behaviors may not occur with sufficient frequency to be understood or dealt with therapeutically, within the course of a given analysis. If the researcher finds that the study of sequences of behavior answers important questions (or poses important problems), then he will devote more effort to improving techniques for studying them.

Another difficulty in the use of behavior sequences for research is that of selecting those sequences which one considers important. Even though psychoanalysts have been making use of patternings of behavior in their daily work for years, research is not oriented toward this sort of data, but rather toward the grouping of people according to classification systems that the researchers themselves claim to be outmoded. New classification schemes based on behavior sequences may eventually be developed if this approach seems fruitful, but in the meantime the selection of problems for research will probably be haphazard and based on personal predilections, just as the selection by the psychotherapist of behavior patterns for interpretation is to some extent based on personal theoretical likes and dislikes.

Finally, how much should be included in a behavior sequence for research? Should it include two contiguous behaviors, three, or how many? And how closely should recurrences of the sequence correspond before they are called identical? We have touched upon these matters in the previous section on research method, but they remain basic problems for psychology.

The Research Method as an Experimental Model

The research model utilizing successive hypotheses has rarely entered into psychological experiments. An experiment within this model would proceed through successive refinements of hypotheses and through successive tests of these hypotheses. *Any* discrepancies from predicted results would lead to a modification of relationships postulated among the variables studied, or to an additional concept. Such an experiment might be conducted with only two subjects, where the statistical test might be the probability of a given ordering among the subjects over a series of experimental situations. Any failure of the ordering would, however, lead to a re-examination of the hypotheses and to the introduction of a new series of experimental situations. From each succeeding series of experiments, the investigator would gain a more precise map of the operant variables. Such a procedure would in some ways resemble a chemical analysis, except that here we are interested in the relationships among variables rather than in the specific subjects, and furthermore, that limitations in our techniques require most often the ordinal handling of data. Aside from its demands on ingenuity in the selection and development of experimental tests as a function of preceding data, such a procedure is crude, and we do not know precisely where it would lead. We are examining our variables and our subjects simultaneously, but if, as we believe, this is true of all psychological investigations, then it is perhaps better that it be done systematically.

SUMMARY

In this paper we have tried to present an analysis of the structural relationships among concepts of psychoanalysis, as posed by the theory of derivatives. We have discussed aspects of the rela-

tionships among impulse, memory trace, anxiety, situation, and defense, and resultant derivative behavior. The final section of this paper considers the methodology from which this theory develops, and some implications of the theory and the methodology for research.

REFERENCES

1. ALLPORT, G. W. The use of personal documents in psychological science. *Soc. Sci. Res. Coun. Bull.*, 1942, No. 49.
2. BIBRING, E. The development and problems of the theory of the instincts. *Int. J. Psychoanal.*, 1941, 22, 102–132.
3. BLUM, G. S. *Psychoanalytic theories of personality.* New York: McGraw-Hill, 1953.
4. ERIKSON, E. H. Observations on the Yurok: childhood and world image. *Univer. Calif. Publ. Amer. Archaeol. Ethnol.*, 1943, 35, 257–302.
5. ERIKSON, E. H. *Childhood and society.* New York: Norton, 1950.
6. FENICHEL, O. *Problems of psychoanalytic technique.* New York: The Psychoanalytic Quarterly, 1941.
7. FENICHEL, O. *The psychoanalytic theory of neurosis.* New York: Norton, 1945.
8. FREUD, ANNA. *The ego and the mechanisms of defence.* New York: International Universities Press, 1946.
9. FREUD, S. *The interpretation of dreams* (1900). Vol. V. *The complete psychological works of Sigmund Freud.* London: Hogarth, 1953.
10. FREUD, S. Recommendations for physicians on the psychoanalytic method of treatment (1912). In *Collected papers.* Vol. II. London: Hogarth, 1949. Pp. 323–334.
11. FREUD, S. Instincts and their vicissitudes (1915). In *Collected papers.* Vol. IV. London: Hogarth, 1949. Pp. 60–84.
12. FREUD, S. Repression (1915). In *Collected papers.* Vol. IV. London: Hogarth, 1949. Pp. 84–98.
13. FREUD, S. Two encyclopaedia articles (1922). In *Collected papers.* Vol. V. London: Hogarth, 1950. Pp. 107–136.
14. FREUD, S. *The problem of anxiety* (1926). New York: Norton, 1936.
15. FREUD, S. *New introductory lectures on psycho-analysis.* New York: Norton, 1933.
16. FREUD, S. Analysis terminable and interminable (1937). In *Collected papers.* Vol. V. London: Hogarth, 1950. Pp. 316–358.
17. HARTMANN, H. Ego psychology and the problem of adaptation (1939). In D. Rapaport (Ed.), *Organization and pathology of thought.* New York: Columbia Univer. Press, 1951. Pp. 362–396.
18. HARTMANN, H. Comments on the psychoanalytic theory of the ego. In *The psychoanalytic study of the child.* Vol. V. New York: International Universities Press, 1950. Pp. 74–97.
19. HARTMANN, H., KRIS, E., & LOEWENSTEIN, R. M. The function of theory in psychoanalysis. In R. M. Loewenstein (Ed.), *Drives, affects, behavior.* New York: International Universities Press, 1953. Pp. 13–38.
20. JONES, E. *The life and work of Sigmund Freud.* Vol. I. New York: Basic Books, 1953.
21. KELLY, E. L., & FISKE, D. W. *The prediction of performance in clinical psychology.* Ann Arbor: Univer. of Michigan Press, 1951.
22. KRIS, E. On preconscious mental processes. *Psychoanal. Quart.*, 1950, 19, 540–560.
23. RAPAPORT, D. *Organization and pathology of thought.* New York: Columbia Univer. Press, 1951.
24. REICH, W. *Character-analysis.* New York: Orgone Institute, 1945.
25. STERBA, R. Introduction to the psychoanalytic theory of the libido. *Nerv. ment. Dis. Monogr.*, 1947, No. 68.

(Received January 27, 1954)

Psychological Review
Vol. 61, No. 6, 1954

A DECISION-MAKING THEORY OF VISUAL DETECTION[1]

WILSON P. TANNER, JR. AND JOHN A. SWETS
University of Michigan

This paper is concerned with the human observer's behavior in detecting light signals in a uniform light background. Detection of these signals depends on information transmitted to cortical centers by way of the visual pathways. An analysis is made of the form of this information, and the types of decisions which can be based on information of this form. Based on this analysis, the expected form of data collected in "yes-no" and "forced-choice" psychophysical experiments is defined, and experiments demonstrating the internal consistency of the theory are presented.

As the theory at first glance appears to be inconsistent with the large quantity of existing data on this subject, it is wise to review the form of these data. The general procedure is to hold signal size, duration, and certain other physical parameters constant, and to observe the way in which the frequency of detection varies as a function of intensity of the light signal. The way in which data of this form are handled implies certain underlying theoretical viewpoints.

In Fig. 1 the dotted lines represent the form of the results of hypothetical experiments. Consider first a single dotted line. Any point on the line might represent an experimentally determined point. This point is corrected for chance by application of the usual formula:

$$p = \frac{p' - c}{1 - c}, \qquad [1]$$

where p' is the observed proportion of positive responses, p is the corrected proportion of positive responses, and c is the intercept of the dotted curve at $\Delta I = 0$.

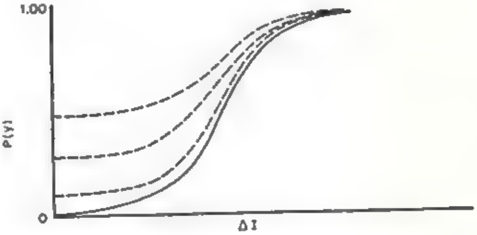

FIG. 1. Conventional seeing frequency or betting curve

Justification of this correction depends on the validity of the assumption that a "false alarm" is a guess, independent of any sensory activity upon which a decision might be based. For this to be the case it is necessary to have a mechanism which triggers when seeing occurs and which becomes incapable of discriminating between quantities of neural activity when seeing does not occur. Only under such a system would a guess be equally likely in the absence of seeing for all values of signal intensity. The application of the chance correction to data from both yes-no and forced-choice experiments is consistent with these assumptions.

The solid curve represents a "true" curve onto which each of the dotted, or experimental, curves can be mapped by using the chance correction and proper estimates of "c." The parameters of the solid curve are assumed to be characteristic of the physiology of the individual's sensory system, independent of psychological control. The assumption carries with it the

[1] This paper is based on work done for the U. S. Army Signal Corps under Contract No. DA-36-039 sc-15358. The experiments reported herein were conducted in the Vision Research Laboratories of the University of Michigan.

notion that if some threshold of neural activity is exceeded, phenomenal seeing results.

To infer that the form of the curve representing the frequency of seeing as a function of light intensity is the same as the curve representing the frequency of seeing as a function of neural activity is to assume a linear relationship between neural activity and light intensity. Efforts to fit seeing frequency curves by normal probability functions suggest a predisposition toward accepting this assumption.

A NEW THEORY OF VISUAL DETECTION

The theory presented in this paper differs from conventional thinking about these assumptions. First, it is assumed that false-alarm rate and correct detection vary together. Secondly, neural activity is assumed to be a monotonically increasing function of light intensity, not necessarily linear. A more specific statement than this is left for experimental determination.

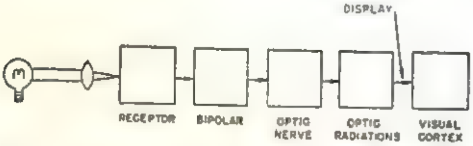

FIG. 2. Block diagram of the visual channel

Figure 2 is a block diagram of the visual pathways showing the major stages of transmission of visual information. All the stages prior to that labelled "cortex" are assumed to function only in the transmission of information, presenting to the cortex a representation of the environment. The function of interpreting this information is left to mechanisms at the cortical level.

In this simplified presentation, the displayed information consists of neural impulse activity. In the case under consideration, in which a signal is presented at a specified time in a known spatial location, the same restrictions are assumed to exist for the display. Thus, if the observer is asked to state whether a signal exists in location *A* at time *B*, he is assumed to consider only that information in the neural display which refers to location *A* at time *B*.

A judgment on the existence of a signal is presumably based on a measure of neural activity. There exists a statistical relationship between the measure and signal intensity. That is, the more intense the signal, the greater is the average of the measures resulting. Thus, for any signal there is a universe distribution which is in fact a sampling distribution. It includes all measures which might result if the signal were repeated and measured an infinite number of times. The mean of this universe distribution is associated with the intensity level of the signal. The variance may be associated with other parameters of the signal such as duration or size, but this is beyond the scope of this paper.

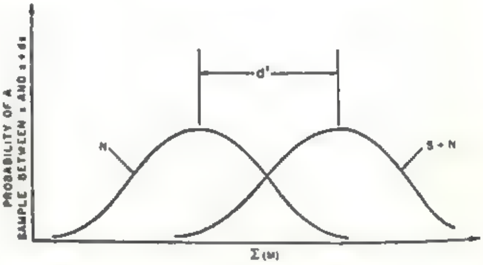

FIG. 3. Hypothetical distributions of noise and signal plus noise

Figure 3 shows two probability distributions: *N* represents the case where noise alone is sampled—that is, no signal exists—and *S* + *N*, the case where signal plus noise exists. The mean of *N* depends on background intensity; the mean of $S + N$ on background plus signal intensity. The variance of N depends on signal parameters, not background parameters in the case considered here; that is, where the observer knows a priori that if a signal exists then it is a particular signal. From the way the diagram is conceptualized, the greater the measure, $\Sigma(M)$, the more likely it is that this sample represents a signal. But one can never be sure. Thus, if an observer is asked if a signal exists, he is assumed to base his judgment on the quantity of neural activity. He makes an observation, and then attempts to decide whether this observation is more representative of N or of $S + N$. His task is, in fact, the task of testing a statistical hypothesis.

The ideal behavior, that which makes optimum use of the information available in this task, is defined mathematically by Peterson and Birdsall (2). The mathematics and symbols used are theirs, unless otherwise stated. The first case considered is the yes-no psychophysical experiment in which a signal is presented at a known location during a well-defined interval in time. This corresponds to Peterson and Birdsall's case of the signal known exactly.

For mathematical convenience, it is assumed that the distributions shown in Fig. 3 are Gaussian, with variance equal for N and all values of $S + N$. Experimental results suggest that equal variance is not a true assumption, but that the deviations are not great enough to justify the inconvenience of a more precise assumption for the purpose of this analysis.

It is also assumed that there is a cutoff point such that any measure of neural activity which exceeds that cutoff is in the criterion; that is, any value exceeding cutoff is accepted as representing the existence of a signal, and any value less than the cutoff represents noise alone. Again, for mathematical convenience, the cutoff point is assumed to be well defined and stable. The justification for accepting this convenience is twofold: first, such behavior is statistically optimum, and second, if absolute stability is physically impossible, any lack of definition or random instability throughout an experiment has the same effect mathematically as additional variance in the sampling distributions.

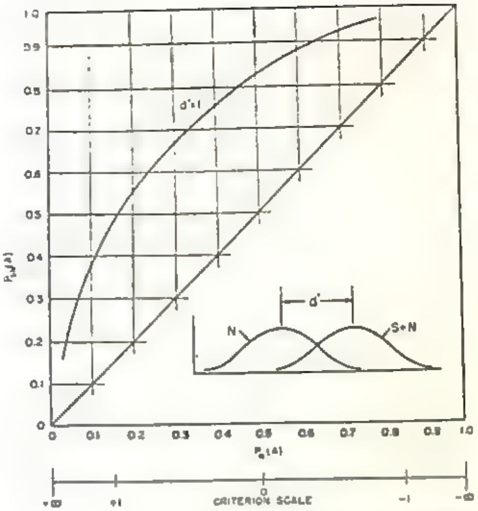

FIG. 4. $P_{SN}(A)$ vs. $P_N(A)$. The criterion scale shows the corresponding criteria expressed in terms of σ_N from M_N.

Now, consider the way in which the placing of the cutoff affects behavior in the case of a given signal. In the lower right-hand corner of Fig. 4 the distributions N and $S + N$ are reproduced for a value of $d' = 1$. The parameter d' is the square root of Peterson and Birdsall's d. The square root of d is more convenient here; d' is the difference between the means of N and $S + N$ in terms of the standard deviation of N. The criterion scale is also calibrated in terms of the standard deviation of N. On the abscissa there is $P_N(A)$, the probability that, if no signal exists, the measure will be in the criterion, and on the ordinate, $P_{SN}(A)$, the probability that if a signal exists, the measure will be in the criterion.

If the cutoff is at $-\infty$, all measures are in the criterion: $P_N(A) = P_{SN}(A) = 1$. At -1 standard deviation, $P_N(A) = .84$, and $P_{SN}(A) = .98$. At 0, $P_N(A) = .5$ and $P_{SN}(A) = .84$. At $+1$, $P_N(A) = .16$ and $P_{SN}(A) = .5$; and for $+\infty$ $P_N(A) = P_{SN}(A) = 0$. Thus, for $d' = 1$ this is the curve showing possible detections for each false-alarm rate. The curve represents the best that can be done with the information available, and the mirror image is the curve of worst possible behaviors.

The maximum behavior in any given experiment is a point on this curve at which the slope is β where

$$\beta = \frac{1-P(SN)}{P(SN)} \frac{(V_{N\cdot CA}+K_{N\cdot A})}{(V_{SN\cdot A}+K_{SN\cdot CA})}. \quad [2]$$

$P(SN)$ is the a priori probability that the signal exists, $V_{N\cdot CA}$ is the value of a correct rejection, $K_{N\cdot A}$ the cost of a false alarm, $V_{SN\cdot A}$ the value of a correct detection, and $K_{SN\cdot CA}$ is the cost of a miss. Thus, as $P(SN)$ or $V_{SN\cdot A}$ increases, or $K_{N\cdot A}$ decreases, β becomes smaller, and it is worth while to accept a higher false-alarm rate in the interest of achieving a greater percentage of correct decisions.

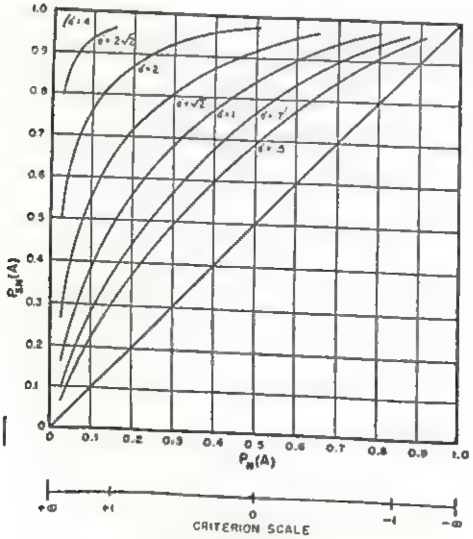

FIG. 5. $P_{SN}(A)$ vs. $P_N(A)$

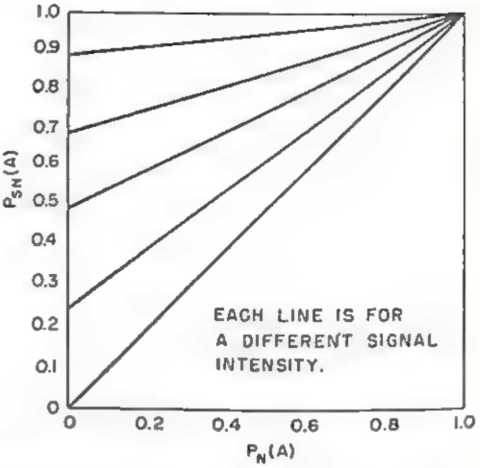

FIG. 6. $P_{SN}(A)$ vs. $P_N(A)$ as a function of d' assuming the guessing hypothesis

Figure 5 shows a family of curves of $P_{SN}(A)$ vs. $P_N(A)$ with d' as a parameter. For values of d' greater than 4, detection is very good. This is to be compared with the predictions of the conventional theory shown in Fig. 6 with $P_N(A)$ assumed to represent guesses. For each value of d' it is assumed that there is a true value of $P_{SN}(A)$ either for $P_N(A) = 0$ or for some very small value. The chance correction should transform each of these to horizontal lines.

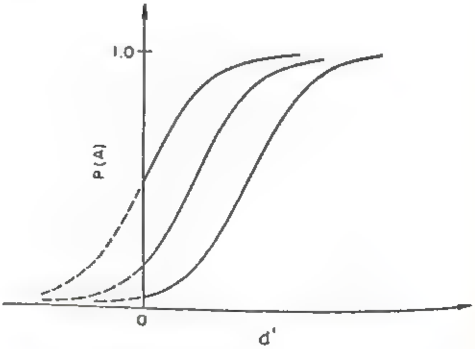

FIG. 7. $P(A)$ as a function of d' assuming the theory

Another way of comparing the predictions of this theory with those of conventional theory is to construct the so-called betting curves, or curves showing the predicted shape of the psychophysical function. These are shown in Fig. 7, where $P(A)$, the probability of acceptance, is plotted as a function of d'. These curves will not map onto the same curve by the application of the chance correction. The shift is horizontal rather than vertical. The dotted portions of the curve show that we are dealing with only a part of the curve, and thus, in the terms of this theory, it is improper to apply a normalizing procedure such as the chance-correction formula to that part of the curve.

In the forced-choice psychophysical experiment, maximum behavior is defined in a different way. In the general forced-choice experiment, the observer knows that the signal will occur in one of n intervals, and he is forced to choose in which of these intervals it occurs. The information upon which his decision is based is contained in the same display as in the case of the yes-no experiment, and, presumably, the values of d' for any given light intensity must be the same. While the solution of this problem is not contained in their study, Peterson and Birdsall have assisted greatly in determining this solution. The probability that a correct answer $P(C)$ will result for a given value of d' is the probability that one sample from the $S + N$ distribution is greater than the greatest of $n - 1$ samples from the distribution of noise alone. The case in which four intervals are used is the basis for Fig. 8. This figure shows the probability of one sample from $S + N$ being greater than the greatest of three from N. For a given value of d' this is

$$P(C) = \int_{x=-\infty}^{+\infty} F(x)^3 g(x) dx, \quad [3]$$

where $F(x)$ is the area of N and $g(x)$ is the ordinate of $S + N$. In Fig. 8 $P(C)$, as determined by this integration, is plotted as a function of d' for the equal-variance case.

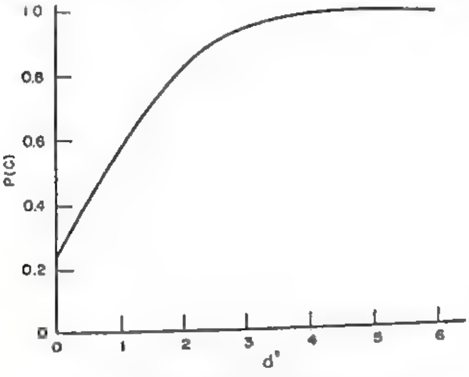

FIG. 8. $P(C)$ as a function of d'. A theoretical curve.

CRITERION OF INTERNAL CONSISTENCY

These two sets of predictions are for the standard experimental situations. They are based on the same neurological parameters. Thus, if the parameters, that is, d's, are estimated from one of the experiments, these estimates should furnish a basis for predicting the data for the other experiment if the theory is internally consistent. An equivalent criterion of internal consistency is that both experiments yield the same estimates of d'.

EXPERIMENTAL DESIGN

Experiments were conducted to test this internal consistency, using three Michigan sophomores as observers. All the experiments employed a circular target 30 minutes in diameter, 1/100 second in duration, on a 10-foot-lambert background. Details of the experimental procedure and the laboratory have been published by Blackwell, Pritchard, and Ohmart (1).

The observers were trained in the temporal forced-choice experiment. The signal appeared in a known location at one of four specified times, and the observers were forced to choose the time at which they thought the signal occurred. Five light intensity increments were used here, with 50 observations per point per experimental session. The last two of these sessions were the test sessions, so that each forced-choice point in the analysis is based on 100 experimental observations.

Following the forced-choice experiments, there was a series of yes-no experiments under the same experimental conditions, except that only four light intensity increments were used. These were the same as the four greatest intensities used in the forced-choice experiments, reduced by adding a .1 fixed filter. In the first four of these sessions, two values of a priori probabilities, $P(SN)$ equal to .8 and .4, were used. The observers were informed of the value of $P(SN)$ before each experimental session. No values or costs were incorporated in these four sessions, which were thus excluded from the analysis as practice sessions.

The test experiments consisted of 12 sessions in each of which all of the information necessary for the calculation of a β (the best best possible decision level) was furnished the observers. While they did not know the formal calculation of β, that they knew the direction of cutoff change indicated by a change in any of these factors was suggested by the fact that the obtained values of $P_N(A)$ varied approximately with changes in the information given them. The values and costs were made real to the observers, for they were actually paid in cash. It was possible for them to earn as much as two dollars extra in a single experimental session as a result of this payment.

The first four sessions each carried the same value of β as $P(SN) = .8$ and the same payment was maintained. A high value of $P_N(A)$, or false-alarm rate, resulted. In the next four sessions with $P(SN)$ held at .8, $K_{N \cdot A}$ and $V_{N \cdot CA}$ were gradually increased from session to session (not within sessions) until $P_N(A)$ dropped to a low value. Then $P(SN)$ was dropped to .4, and K_{NA} and $V_{N \cdot CA}$ were reduced so that for the thirteenth session $P_N(A)$ stayed low. The last three sessions successively involved increases in $V_{S \cdot NA}$ and $K_{SN \cdot CA}$, again forcing $P_N(A)$ toward a higher value.

RESULTS

Figures 9 and 10 show scatter diagrams of $P_{SN}(A)$ vs. $P_N(A)$ for a particular intensity of signal and for a single observer. These scatter diagrams can be used to estimate d'. In Fig. 9 the estimate of d' is .7. In Fig. 10, the estimate of d' is 1.3. Each d' estimated in this way is based on 560 observations. A procedure similar to this was used for the d's for each of four signals for each of the four observers.

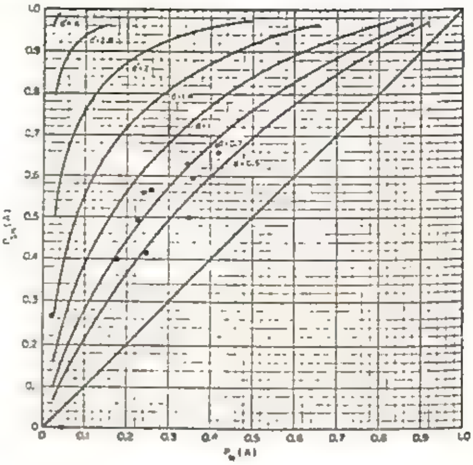

FIG. 9. A scatter diagram of $P_{SN}(A)$ vs. $P_N(A)$

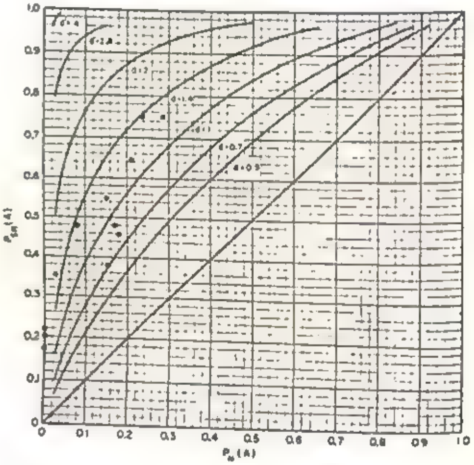

FIG. 10. A scatter diagram of $P_{SN}(A)$ vs. $P_N(A)$

In the forced-choice experiment the estimates of d' are made by entering our forced-choice curve (Fig. 8), using the observed percentage correct as an estimate of $P(C)$. Figure 11 shows log d' as a function of log signal in-

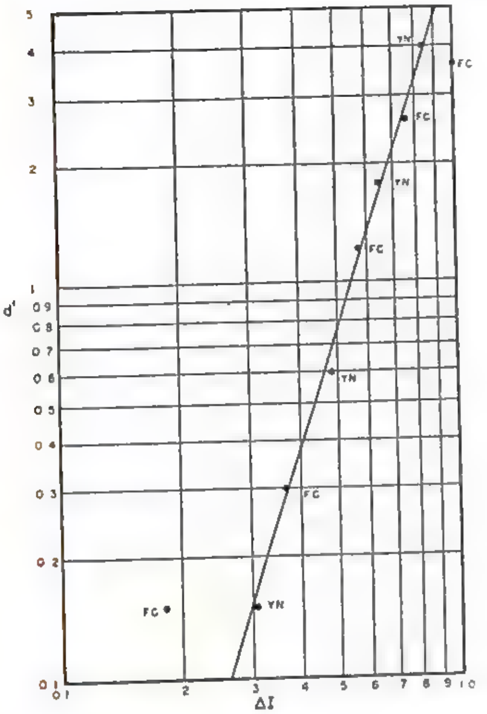

FIG. 11. Log *d'* vs. Log ΔI for Observer 1

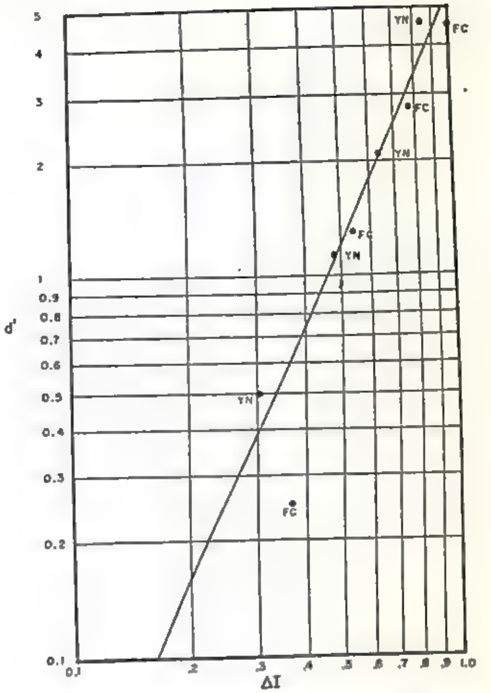

FIG. 13. Log *d'* vs. Log ΔI for Observer 3

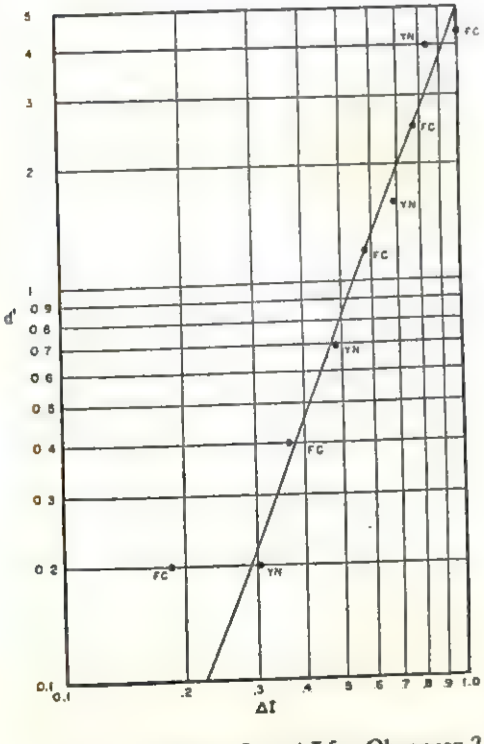

FIG. 12. Log *d'* vs. Log ΔI for Observer 2

tensity for the first observer, the estimates of *d'* being from both forced-choice and yes-no experiments. In general the agreement is good. The deviation of the forced-choice point at the top can be explained on the basis of inadequate experimental data for the determination of the high probability involved. The deviation of the low point is unexplained. Figure 12 is the same plot for the second observer, showing about the same picture. Figure 13 is for the third observer, showing not quite as good a fit, but nevertheless satisfactory for psychological experiments. For this observer, the lowest point for forced choice is off the graph to the right of the line.

Figures 14, 15, and 16 show the predictions for forced-choice data (when yes-no data are used to estimate *d'*) for the three observers. Note that the lowest point is on the curve in both of the first two cases,

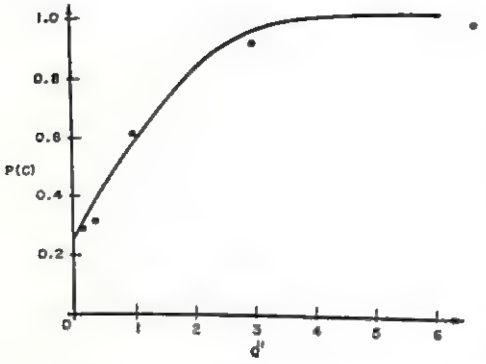

Fig. 14. Prediction of forced-choice data from yes-no data for Observer 1

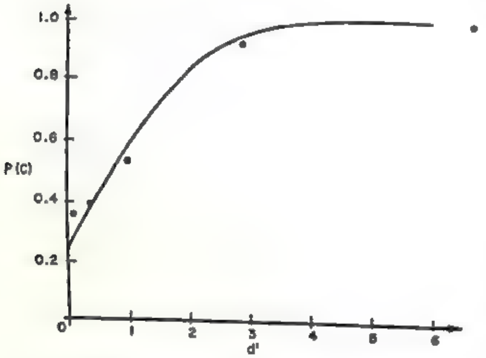

Fig. 15. Prediction of forced-choice data from yes-no data for Observer 2

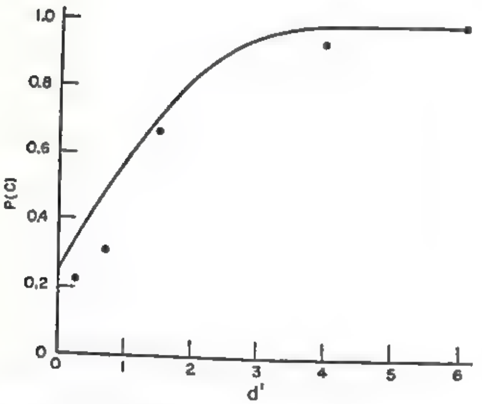

Fig. 16. Prediction of forced-choice data from yes-no data for Observer 3

suggesting that the deviation which appeared on the curves in Fig. 11, 12, and 13 is not significant.

Discussion

The results satisfy the criterion of internal consistency. The theory also turns out to be consistent with the vast amount of data in the literature, for, when the d' vs. ΔI function for any one of the observers is used to predict probability of detection as a function of ΔI in terms of this theory, the result closely approximates the type of curve frequently reported. Shapes of curves thus furnish no basis for selecting between the two theories, and a decision must rest on the other arguments.

According to conventional theory, application of the chance correction should yield corrected values of $P_{SN}(A)$ which are independent of $P_N(A)$, or should yield corrected thresholds in the conventional sense which are independent of $P_N(A)$. Rank-order correlations for the three observers between $P_N(A)$ and corrected thresholds (.30, .71, .67) are highly significant; the combined $p \ll .001$. This is a result consistent with theory presented here.

Another method of comparison is to fit the scatter diagrams (Fig. 9 and 10) by straight lines. According to the independence theory, these straight lines should intercept the point (1.00, 1.00). Sampling error would be expected to send some of the lines to either side of this point. There are 12 of these scatter diagrams, and all 12 of these lines intersect the line $P_{SN}(A) = 1.00$ at values of $P_N(A)$ between 0 and 1.00 in an order which would be predicted if these lines were arcs of the curves $P_{SN}(A)$ vs. $P_N(A)$ as defined by the theory of signal detectability.

Two additional sessions were run in which the observers were permitted three categories of response (yes, no, and doubtful), and were told to be sure of being correct if they responded either yes or no. Again, two a priori probabilities (.8 and .4) were employed, and again $P_N(A)$ was correlated with $P(SN)$. The observers, interviewed after these sessions, reported that their "yes" responses were based on "phenomenal" seeing.

This does not mean that the observers were abnormal because they hallucinated. It suggests, on the other hand, that phenomenal seeing develops through experience, and is subject to change with experience. Psychological as well as physiological factors are involved. Psychological "set" is a function of β, and after experience with a given set one begins to see, or not to see, rather automatically. Change the set, and the level of seeing changes. The experiments reported here were such that the observers learned to adjust rapidly to different sets.

CONCLUSIONS

The following conclusions are advanced: (*a*) The conventional concept of a threshold, or a threshold region, needs re-evaluation in the light of the present theory that the visual detection problem is the problem of detecting signals in noise. (*b*) The hypothesis that false alarms are guesses is rejected on the basis of statistical tests. (*c*) Change in neural activity is a power function of change in light intensity. (*d*) The mathematical model of signal detection is applicable to problems of visual detection. (*e*) The criterion of seeing depends on psychological as well as physiological factors. In the experiments reported here the observers tended to use optimum criteria. (*f*) The experimental data support the assumption of a logical connection between forced-choice and yes-no techniques developed by the theory.

REFERENCES

1. BLACKWELL, H. R., PRITCHARD, B. S., & OHMART, T. G. Automatic apparatus for stimulus presentation and recording in visual threshold experiments. *J. opt. Soc. Amer.*, 1954, 44, 322–326.
2. PETERSON, W. W., & BIRDSALL, T. G. The theory of signal detectability. Electronic Defense Group, Univer. of Michigan, *Tech. Rep.*, No. 13, Sept., 1953.

(Received October 29, 1953)

Psychological Review
Vol. 61, No. 6, 1954

ON ABELSON'S CRITICAL COMMENT

DAVID BAKAN

University of Missouri

Abelson's paper (1) leads me to conclude that either the original paper (2) is very unclear, or Abelson misunderstood it, or both. Under any circumstances, given what seems to be his understanding of it, he has indeed been very gracious in his criticism; for that which he attributes to the paper makes it patently absurd.

Essentially he has ascribed his own definitions, and at least one additional assumption, to the original development and, very understandably, finds that this leads to difficulties.

In the original paper $P(g)$ is used in the sense of the *strength* of g, in much the same way as Thorndike thought of the strength of a connection (2, p. 368). If g is the "what is learned" (2, p. 362), $P(g)$ is the degree of it. It appears likely, particularly in view of Abelson's misinterpretation, that this was perhaps not sufficiently explicated in the original paper.

Abelson defines $P(g)$ as "the probability that the organism is in the condition g" (1, p. 276). It is difficult to conceive of a definition more at variance with both the intent and the content of the original paper. By his definition the locus of $P(g)$ is *outside* the organism. He is distinctly in error when he attributes an outside-the-organism $P(g)$, in this simple sense, to "Bakan's model" (1, p. 278). The outside-the-organism $P(g)$ stems rather from *his* definition of $P(g)$.

A similar straw man which he sets up for criticism is contained in his phrase, "If we take seriously Bakan's formulation, which allows only two conditions for the organism, g and 'not-g' . . ." (1, p. 277). It is certainly true that the paper centers around g and $\bar{g}$, but no one has ever concluded that there are only two kinds of animals because the animal universe may be divided into horses and not-horses. No two-g assumption is made in the paper at all. The conclusions that Abelson derives based on this assumption are interesting, but they are his own.

REFERENCES

1. Abelson, R. B. Critical comment on "Learning and the principle of inverse probability." *Psychol. Rev.*, 1954, **61**, 276–278.
2. Bakan, D. Learning and the principle of inverse probability. *Psychol. Rev.*, 1953, **60**, 360–370.

(Received July 6, 1954)

PSYCHOLOGICAL REVIEW

July 1, 1954

YEAR	VOLUME	AVAILABLE NUMBERS						PRICE PER NUMBER	PRICE PER VOLUME
1894	1	–	2	–	4	5	6	$1.50	——
1895	2	–	–	3	4	5	6	$1.50	——
1896	3	–	–	–	–	–	–	——	——
1897	4	1	–	–	–	–	6	$1.50	——
1898	5	–	–	3	–	5	–	$1.50	——
1899	6	–	–	–	–	–	–	$1.50	——
1900	7	1	–	–	–	–	–	$1.50	——
1901	8	1	2	–	–	–	–	$1.50	——
1902	9	–	2	–	–	–	–	$1.50	——
1903	10	1	2	–	–	–	–	$1.50	——
1904	11	1	–	–	4 &	5	6	$1.50	——
1905	12	1	2 &	3	4	5	–	$1.50	——
1906	13	–	–	3	4	5	6	$1.50	——
1907	14	1	2	–	–	–	–	$1.50	——
1908	15	–	–	–	–	–	–	——	——
1909	16	1	–	3	4	5	–	$1.50	——
1910	17	1	2	3	–	–	6	$1.50	——
1911	18	1	2	3	4	5	–	$1.50	——
1912	19	1	2	3	4	5	6	$1.50	$8.00
1913	20	1	2	3	4	5	6	$1.50	$8.00
1914	21	–	2	3	4	5	6	$1.50	——
1915	22	1	2	3	4	5	6	$1.50	$8.00
1916	23	1	–	–	4	5	–	$1.50	——
1917	24	–	–	–	4	–	–	$1.50	——
1918	25	–	2	3	4	5	6	$1.50	——
1919	26	1	2	3	4	5	6	$1.50	$8.00
1920	27	1	–	–	–	–	–	$1.50	——
1921	28	–	2	–	–	–	6	$1.50	——
1922	29	1	–	–	4	–	–	$1.50	——
1923	30	1	–	3	–	–	–	$1.50	——
1924	31	1	2	3	4	5	6	$1.50	$8.00
1925	32	–	2	3	–	5	–	$1.50	——
1926	33	1	–	3	4	5	6	$1.50	——
1927	34	1	2	3	4	5	6	$1.50	$8.00
1928	35	1	2	3	4	5	6	$1.50	$8.00
1929	36	1	2	3	4	5	6	$1.50	$8.00
1930	37	1	2	3	4	5	6	$1.50	$8.00
1931	38	1	2	3	4	5	6	$1.50	$8.00
1932	39	1	2	3	4	5	6	$1.50	$8.00
1933	40	1	2	3	4	5	6	$1.50	$8.00
1934	41	1	2	3	4	5	6	$1.50	$8.00
1935	42	1	2	3	4	5	6	$1.50	$8.00
1936	43	1	2	3	4	5	6	$1.50	$8.00
1937	44	1	2	3	4	5	6	$1.50	$8.00
1938	45	1	2	3	4	5	6	$1.50	$8.00
1939	46	1	2	3	4	5	6	$1.50	$8.00
1940	47	–	–	3	4	–	–	$1.50	——
1941	48	–	–	–	–	–	6	$1.50	——
1942	49	–	2	3	4	5	6	$1.50	——
1943	50	1	2	3	4	5	6	$1.50	$8.00
1944	51	1	2	3	4	5	6	$1.50	$8.00
1945	52	1	2	3	4	–	6	$1.50	——
1946	53	1	2	3	4	5	6	$1.50	$8.00
1947	54	1	2	–	4	5	6	$1.50	——
1948	55	1	2	3	4	5	6	$1.50	$8.00
1949	56	1	2	3	4	5	6	$1.50	$8.00
1950	57	1	2	3	4	5	6	$1.50	$8.00
1951	58	1	2	3	4	5	6	$1.50	$8.00
1952	59	1	2	3	4	5	6	$1.50	$8.00
1953	60	1	2	3	4	5	6	$1.50	$8.00
1954	61	By subscription $6.50, foreign $7.00						$1.50	——

Postage prepaid on U. S. orders. Add $.50 per volume on foreign orders. All stock subject to prior sale. The American Psychological Association gives the following discounts on orders for any one journal:

10% on orders of $ 50.00 and over
20% on orders of $100.00 and over
30% on orders of $150.00 and over

Current subscriptions and orders for back numbers should be addressed to

AMERICAN PSYCHOLOGICAL ASSOCIATION
1333 Sixteenth Street N.W.
Washington 6, D. C.

PSYCHOLOGICAL MONOGRAPH

#270

1945

ON PROBLEM SOLVING

By

KARL DUNCKER

$2.50

This popular monograph is #270 of the Psychological Monograph series. It has been reprinted so that it is again available.

Third Printing

American Psychological Association
1333 Sixteenth Street N. W.
Washington 6, D. C.

✓